National Institute for Health and Care Excellence

Diabetes in pregnancy: management from preconception to the postnatal period

NICE guideline NG3

Methods

June 2026

Draft for consultation

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or their carer or guardian.

Local commissioners and/or providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

ISBN: [XXX]

Contents

Development of the guideline

What this guideline covers

This guideline update covers recommendations and research recommendations on managing type 1 diabetes using hybrid closed loop systems in people who are planning to become pregnant, are pregnant, or are in the postpartum period.

What this guideline does not cover

All other recommendations in this guideline were developed using the methods outlined in either the evidence review documents associated with the December 2020 update, or the methods chapter in the 2015 full guideline, both available here: Overview | Diabetes in pregnancy: management from preconception to the postnatal period | Guidance | NICE

Methods

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual.

Declarations of interest were recorded according to NICE's conflicts of interest policy.

Developing the review question and outcomes

One review question was developed for this guideline which was based on the key areas identified in the guideline scope. This was drafted by the NICE guideline development team and refined and validated by the guideline committee.

The review question was based on the following framework:

- Population, Intervention, Comparator and Outcome and study type (PICOS) for review of intervention

Full literature searches, critical appraisals and evidence review were completed for the review question.

Reviewing research evidence

Review protocol

The review protocol was developed with the guideline committee to outline the inclusion and exclusion criteria used to select studies for the evidence review. The review protocol was prospectively published on the NICE website.

Searching for evidence

Evidence was searched for the review question using the methods specified in Developing NICE guidelines: the manual.

Selecting studies for inclusion

All references identified by the literature searches and from other sources (for example, previous versions of the guideline or studies identified by committee members) were uploaded into EPPI reviewer software (version 5) and de-duplicated. Titles and abstracts were assessed for possible inclusion using the criteria specified in the review protocol. At least 10% of the abstracts were reviewed by two reviewers, with any disagreements resolved by discussion or, if necessary, a third independent reviewer.

The full text of potentially eligible studies was retrieved and assessed according to the criteria specified in the review protocol. A standardised form was used to extract data from included studies. Study investigators were contacted for missing data when time and resources allowed (when this occurred, this was noted in the evidence review and relevant data was included).

Incorporating published evidence syntheses

If published evidence syntheses were identified sufficiently early in the review process (for example, from the surveillance review or early in the database search), they were considered for use as the primary source of data, rather than extracting information from primary studies. Syntheses considered for inclusion in this way were quality assessed to assess their suitability using the appropriate checklist, as outlined in **Error! Reference source not found.**. Note that this quality assessment was solely used to assess the quality of the synthesis in order to decide whether it could be

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

Table 1: Checklists for published evidence syntheses

Type of synthesis	Checklist for quality appraisal
Systematic review of quantitative evidence	ROBIS

Each published evidence synthesis was classified into one of the following three groups:

- High quality – It is unlikely that additional relevant and important data would be identified from primary studies compared to that reported in the review, and unlikely that any relevant and important studies have been missed by the review.
- Moderate quality – It is possible that additional relevant and important data would be identified from primary studies compared to that reported in the review, but unlikely that any relevant and important studies have been missed by the review.
- Low quality – It is possible that relevant and important studies have been missed by the review.

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

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

The way that a published evidence synthesis was used in the evidence review depended on its quality and applicability, as defined in Table 2. When published

evidence syntheses were used as a source of primary data, data from these evidence syntheses were quality assessed and presented in GRADE tables in the same way as if data had been extracted from primary studies. In questions where data was extracted from both systematic reviews and primary studies, these were checked to ensure none of the data had been double counted through this process.

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

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

Methods of combining evidence

Data synthesis for intervention studies

Where possible, meta-analyses were conducted to combine the results of quantitative studies for each outcome.

Pairwise meta-analysis

Pairwise meta-analyses were performed in Cochrane RevMan Web. A pooled relative risk was calculated for dichotomous outcomes (using the Mantel–Haenszel method) reporting numbers of people having an event. When studies reporting dichotomous outcomes had zero events in one arm (single-arm-zero-event studies) and combined event rates in both intervention and control arms were below 1%, Peto odds ratio (POR) was used as a summary measure. For dichotomous outcomes with zero events in both arms (double-arm-zero-event studies), where some or all studies reported no events in either group (intervention or control), data was analysed using risk difference (RD) as an appropriate summary measure.

A pooled mean difference was calculated for continuous outcomes (using the inverse variance method) when the same scale was used to measure an outcome across different studies.

For continuous outcomes analysed as mean differences, final timepoint values were used in the meta-analysis if they were accompanied by a measure of spread (for example standard deviation). When multiple final time points were available for a specific period (for example, by trimester), and no overall combined estimate was provided, the time-specific data were assessed for variability. If the results showed minimal differences between time points, the review team extracted either the dataset with the largest number of participants or the dataset judged to be the most representative. When a study provided both time-specific estimates and an overall pooled result, the overall estimate was extracted in accordance with NICE's preference for the least selective and most comprehensive summary.

Fixed-effects models were the preferred choice to report, but in situations where the assumption of a shared mean for fixed-effects model were clearly not met, even after appropriate pre-specified subgroup analyses were conducted, random-effects results

are presented. Fixed-effects models were deemed to be inappropriate if there was significant statistical heterogeneity in the meta-analysis, defined as I^2 more than or equal to 50%.

However, in cases where the results from individual pre-specified subgroup analyses were less heterogeneous (with I^2 less than 50%) the results from these subgroups were reported using fixed effects models. This may have led to situations where pooled results were reported from random-effects models and subgroup results were reported from fixed-effects models.

Where sufficient studies were available, meta-regression was considered to explore the effect of study level covariates.

Appraising the quality of evidence

Intervention studies (relative effect estimates)

RCTs were quality assessed using the Cochrane Risk of Bias Tool (version 2.0). Evidence on each outcome for each individual study was classified into one of the following groups:

- Low risk of bias – The true effect size for the study is likely to be close to the estimated effect size.
- Moderate risk of bias – There is a possibility the true effect size for the study is substantially different to the estimated effect size.
- High risk of bias – It is likely the true effect size for the study is substantially different to the estimated effect size.

.

Each individual study was also classified into one of three groups for directness, based on if there were concerns about the population, intervention, comparator and/or outcomes in the study and how directly these variables could address the specified review question. Studies were rated as follows:

- Direct – No important deviations from the protocol in population, intervention, comparator and/or outcomes.
- Partially indirect – Important deviations from the protocol in one of the following areas: population, intervention, comparator and/or outcomes.
- Indirect – Important deviations from the protocol in at least two of the following areas: population, intervention, comparator and/or outcomes.

Minimally important differences (MIDs) and clinical decision thresholds

The Core Outcome Measures in Effectiveness Trials (COMET) database was searched to identify published minimal clinically important difference thresholds relevant to this guideline that might aid the committee in identifying clinical decision thresholds for the purpose of GRADE. Identified MIDs were assessed to ensure they had been developed and validated in a methodologically rigorous way, and were applicable to the populations, interventions and outcomes specified in this guideline. In addition, the Guideline Committee were asked to prospectively specify any

outcomes where they felt a consensus clinical decision threshold could be defined from their experience. In particular, any questions looking to evaluate non-inferiority (that one treatment is not meaningfully worse than another) required a clinical decision threshold to be defined to act as a non-inferiority margin.

Clinical decision thresholds were used to assess imprecision using GRADE and aid interpretation of the size of effects for different outcomes. Clinical decision threshold that were used in the guideline are reported in the relevant evidence reviews.

GRADE for intervention studies analysed using pairwise analysis

GRADE was used to assess the quality of evidence for the outcomes specified in the review protocol. Data from randomised controlled trials (which were quality assessed using the Cochrane risk of bias tool) were initially rated as high quality. The quality of the evidence for each outcome was downgraded or not from this initial point, based on the criteria given in Table 3. These criteria were used to apply preliminary ratings, but were overridden in cases where, in the view of the analyst or committee the uncertainty identified was unlikely to have a meaningful impact on decision making.

Table 3: Rationale for downgrading quality of evidence for intervention studies

GRADE criteria	Reasons for downgrading quality
Risk of bias	• Not serious (don't downgrade): less than 50% overall weighting some concerns/high risk of bias • Serious (downgrade 1 level): more than 50% some concerns/high risk of bias • Very serious (downgrade 2 levels): more than 50% high risk of bias.
Indirectness	• Not serious (don't downgrade): less than 50% of overall weighting partially direct or indirect. • Serious (downgrade 1 level): more than 50% of overall weighting partially direct or indirect. • Very serious (downgrade 2 levels): more than 50% of overall weighting indirect
Inconsistency	Concerns about inconsistency of effects across studies, occurring when there is unexplained variability in the treatment effect demonstrated across studies (heterogeneity), after appropriate pre-specified subgroup analyses have been conducted. This was assessed using the I^2 statistic. • Not serious (don't downgrade) I^2 = less than 40%; • Serious (downgrade 1 level) I^2 = 40-60%; • Very serious (downgrade 2 levels) I^2 = more than 60%.

GRADE criteria	Reasons for downgrading quality
Imprecision	Where established MIDs are available these were used. If there are no established MIDs imprecision was assessed using Optimal Information Size and event rates. Interpretation of imprecision and consideration of baseline event rates was discussed with the committee whenever feasible and a summary included in the CDE section of each evidence review. If an MID other than the line of no effect was defined for the outcome, the outcome was downgraded once if the 95% confidence interval for the effect size crossed one line of the MID, and twice if it crosses both lines of the MID. If the line of no effect was defined as an MID for the outcome, it was downgraded once if the 95% confidence interval for the effect size crossed the line of no effect (i.e. the outcome was not statistically significant), and twice if the sample size of the study was sufficiently small that it is not plausible any realistic effect size could have been detected. Outcomes meeting the criteria for downgrading above were not downgraded if the confidence interval was sufficiently narrow that the upper and lower bounds would correspond to clinically equivalent scenarios. If Optimal Information Size was used, outcomes were downgraded twice if the sample size was smaller than 30% of the Optimal Information Size.
Publication bias	Where 10 or more studies were included as part of a single meta-analysis, a funnel plot was produced to graphically assess the potential for publication bias. When a funnel plot showed convincing evidence of publication bias, or the review team became aware of other evidence of publication bias (for example, evidence of unpublished trials where there was evidence that the effect estimate differed in published and unpublished data), the outcome was downgraded once. If no evidence of publication bias was found for any outcomes in a review (as was often the case), this domain was excluded from GRADE profiles to improve readability.

Reviewing economic evidence

Identifying economic evidence

A systematic review of economic literature was conducted for the 1 review question covered in the guideline update. Titles and abstracts of articles identified through the systematic economic literature searches were assessed for inclusion using predefined eligibility criteria reported in the economic review protocol (provided in the technical appendix A to the evidence review).

Once the screening of titles and abstracts was completed, full-text copies of potentially relevant articles were acquired for detailed assessment, applying the economic review protocol inclusion and exclusion criteria.

Details of economic evidence study selection and lists of excluded studies after full-text assessment (together with reasons for exclusion) are presented in respective appendices to the evidence review.

Appraising the quality of economic evidence

The applicability and methodological quality of economic evidence derived either from published studies meeting the inclusion criteria or from new economic analysis conducted for the guideline was assessed using the economic evaluations checklist specified in Developing NICE guidelines: the manual, Appendix H. This process led to applicability and quality statements for each included study, made by the health economist, following the criteria shown in Table 4.

Table 4: Criteria for developing applicability and quality statements of economic evidence

Appraised element	Statement and criteria
Applicability	• Directly applicable – the study meets all applicability criteria, or fails to meet 1 or more applicability criteria but this is unlikely to change the conclusions about cost effectiveness. • Partially applicable – the study fails to meet 1 or more applicability criteria, and this could change the conclusions about cost effectiveness. • Not applicable – the study fails to meet 1 or more of the applicability criteria, and this is likely to change the conclusions about cost effectiveness. Such studies would usually be excluded from the review.

Appraised element	Statement and criteria
Quality	• Minor limitations – the study meets all quality criteria, or fails to meet 1 or more quality criteria, but this is unlikely to change the conclusions about cost effectiveness. • Potentially serious limitations – the study fails to meet 1 or more quality criteria, and this could change the conclusions about cost effectiveness. • Very serious limitations – the study fails to meet 1 or more quality criteria, and this is highly likely to change the conclusions about cost effectiveness. Such studies would usually be excluded from the review.

The single study that partially met the applicability and quality criteria described in the methodology checklist were considered by the committee during the guideline development process. Further criteria for prioritising economic studies for use in decision making are provided in the economic review protocols (under 'Review strategy'), within the respective appendices of the evidence reviews.

Details on methods and results of the economic study that met the inclusion criteria and was subsequently used in decision making are shown in economic evidence study extraction tables, provided in respective appendices of the technical appendix to the evidence review.

Characteristics and results (cost-effectiveness estimates) of economic studies used in decision making, including applicability and quality statements, have been summarised in economic evidence characteristics and summary tables, respectively, provided in the economic sections of the evidence review.

New economic analysis

No economic modelling was undertaken for this guideline update which required the incorporation of Technology Appraisal 943.

Cost effectiveness criteria

NICE's principles set out criteria 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 overall resource use and more effective compared with all other relevant alternative strategies)
- the intervention cost less than £20,000 per QALY gained compared with the next best strategy.

If the committee recommended an intervention that was estimated to cost more than £20,000 per QALY gained, or did not recommend one that was estimated to cost less than £20,000 per QALY gained, then the reasons for this decision were provided and recorded, with reference to issues around the plausibility of the estimate or to other factors, for example the degree of uncertainty around the ICER, aspects that relate to uncaptured benefits and non-health factors, or aspects that relate to health inequalities, as set out in the NICE health technology evaluations manual.

The committee's considerations of cost effectiveness are discussed explicitly in the section 'Committee discussion and interpretation of the evidence' under the subheading 'Resources and cost-effectiveness', in the evidence review.

Glossary of terms

Abstract: Summary of a study, which may be published alone or as an introduction to a full scientific paper.

Applicability: How well the results of a study or NICE evidence review can answer a clinical question or be applied to the population being considered.

Arm: Subsection of individuals within a study who receive one particular intervention, for example placebo arm.

Association: Statistical relationship between 2 or more events, characteristics or other variables. The relationship may or may not be causal.

Baseline: The initial set of measurements at the beginning of a study (after run-in period where applicable), with which subsequent results are compared.

Bias: Influences on a study that can make the results look better or worse than they really are. (Bias can even make it look as if a treatment works when it does not.) Bias can occur by chance, deliberately or as a result of systematic errors in the design and execution of a study. It can also occur at different stages in the research process, for example, during the collection, analysis, interpretation, publication or review of research data. For examples see selection bias, performance bias, information bias, confounding factor, and publication bias.

Carer (caregiver): Someone who looks after family, partners or friends in need of help because they are ill, frail or have a disability.

Cochrane: The Cochrane Library consists of a regularly updated collection of evidence-based medicine databases including the Cochrane Database of Systematic Reviews (reviews of randomised controlled trials prepared by the Cochrane Collaboration).

Confidence interval (CI): A range of values for an unknown population parameter with a stated ‘confidence’ (conventionally 95%) that it contains the true value. The interval is calculated from sample data, and generally straddles the sample estimate. The ‘confidence’ value means that if the method used to calculate the interval is repeated many times, then that proportion of intervals will actually contain the true value.

Confounding factor: Something that influences a study and can result in misleading findings if it is not understood or appropriately dealt with. For example, a study of heart disease may look at a group of people that exercises regularly and a group that does not exercise. If the ages of the people in the 2 groups are different, then any difference in heart disease rates between the 2 groups could be because of age rather than exercise. Therefore, age is a confounding factor.

Control group: A group of people in a study who do not receive the treatment or test being studied. Instead, they may receive the standard treatment (sometimes called 'usual care') or a dummy treatment (placebo). The results for the control group are compared with those for a group receiving the treatment being tested. The aim is to check for any differences. Ideally, the people in the control group should be as similar as possible to those in the treatment group, to make it as easy as possible to detect any effects due to the treatment.

Cost-effectiveness analysis (CEA): Cost-effectiveness analysis is one of the tools used to carry out an economic evaluation. The benefits are expressed in non-monetary terms related to health, such as symptom-free days, heart attacks avoided, deaths avoided or life years gained (that is, the number of years by which life is extended as a result of the intervention).

Cost-effectiveness model: An explicit mathematical framework, which is used to represent clinical decision problems and incorporate evidence from a variety of sources in order to estimate the costs and health outcomes.

Economic evaluation: An economic evaluation is used to assess the cost effectiveness of healthcare interventions (that is, to compare the costs and benefits of a healthcare intervention to assess whether it is worth doing). The aim of an economic evaluation is to maximise the level of benefits – health effects – relative to the resources available. It should be used to inform and support the decision-making process; it is not supposed to replace the judgement of healthcare professionals. There are several types of economic evaluation: cost–benefit analysis, cost–consequences analysis, cost-effectiveness analysis, cost minimisation analysis and cost–utility analysis. They use similar methods to define and evaluate costs, but differ in the way they estimate the benefits of a particular drug, programme or intervention.

Effect: A measure that shows the magnitude of the outcome in one group compared with that in a control group. For example, if the absolute risk reduction is shown to be 5% and it is the outcome of interest, the effect size is 5%. The effect size is usually tested, using statistics, to find out how likely it is that the effect is a result of the treatment and has not just happened by chance (that is, to see if it is statistically significant).

Effectiveness: How beneficial a test or treatment is under usual or everyday conditions, compared with doing nothing or opting for another type of care.

Evidence: Information on which a decision or guidance is based. Evidence is obtained from a range of sources including randomised controlled trials, observational studies, expert opinion (of clinical professionals or patients).

Exclusion criteria: Explicit standards used to decide which studies should be excluded from consideration as potential sources of evidence.

GRADE: A system developed by the GRADE Working Group to address the shortcomings of present grading systems in healthcare. The GRADE system uses a common, sensible and transparent approach to grading the quality of evidence. The results of applying the GRADE system to clinical trial data are displayed in a table known as a GRADE evidence profile.

Heterogeneity: The term is used in meta-analyses and systematic reviews to describe when the results of a test or treatment (or estimates of its effect) differ significantly in different studies. Such differences may occur as a result of differences in the populations studied, the outcome measures used or because of different definitions of the variables involved. It is the opposite of homogeneity.

Imprecision: Results are imprecise when studies include relatively few patients and few events and thus have wide confidence intervals around the estimate of effect.

Inclusion criteria: Explicit criteria used to decide which studies should be considered as potential sources of evidence.

Incremental cost effectiveness ratio (ICER): The difference in the mean costs in the population of interest divided by the differences in the mean outcomes in the population of interest for one treatment compared with another.

Indirectness: The available evidence is different to the review question being addressed, in terms of PICO (population, intervention, comparison and outcome).

Intervention: In medical terms this could be a drug treatment, surgical procedure, diagnostic or psychological therapy. Examples of public health interventions could include action to help someone to be physically active or to eat a healthier diet.

Life years gained: Mean average years of life gained per person as a result of the intervention compared with an alternative intervention.

Meta-analysis: A method often used in systematic reviews. Results from several studies of the same test or treatment are combined to estimate the overall effect of the treatment.

Outcome: The impact that a test, treatment, policy, programme or other intervention has on a person, group or population. Outcomes from interventions to improve the public's health could include changes in knowledge and behaviour related to health, societal changes (for example, a reduction in crime rates) and a change in people's health and wellbeing or health status. In clinical terms, outcomes could include the number of patients who fully recover from an illness or the number of hospital admissions, and an improvement or deterioration in someone's health, functional ability, symptoms or situation. Researchers should decide what outcomes to measure before a study begins.

Publication bias: Publication bias occurs when researchers publish the results of studies showing that a treatment works well and don't publish those showing it did not have any effect. If this happens, analysis of the published results will not give an accurate idea of how well the treatment works. This type of bias can be assessed by a funnel plot.

Quality-adjusted life year (QALY): A measure of the state of health of a person or group in which the benefits, in terms of length of life, are adjusted to reflect the quality of life. One QALY is equal to 1 year of life in perfect health. QALYS are calculated by estimating the years of life remaining for a patient following a particular treatment or intervention and weighting each year with a quality of life score (on a

scale of 0 to 1). It is often measured in terms of the person's ability to perform the activities of daily life, freedom from pain and mental disturbance.

Randomised controlled trial (RCT): A study in which a number of similar people are randomly assigned to 2 (or more) groups to test a specific drug or treatment. One group (the experimental group) receives the treatment being tested, the other (the comparison or control group) receives an alternative treatment, a dummy treatment (placebo) or no treatment at all. The groups are followed up to see how effective the experimental treatment was. Outcomes are measured at specific times and any difference in response between the groups is assessed statistically. This method is also used to reduce bias.

Review question: In guideline development, this term refers to the questions about treatment and care that are formulated to guide the development of evidence-based recommendations.

Sub-groups: Planned statistical investigations if heterogeneity is found in the meta-analysis. Specified a priori in the review protocol.

Systematic review: A review in which evidence from scientific studies has been identified, appraised and synthesised in a methodical way according to predetermined criteria. It may include a meta-analysis.

Utility: In health economics, a 'utility' is the measure of the preference or value that an individual or society places upon a particular health state. It is generally a number between 0 (representing death) and 1 (perfect health). The most widely used measure of benefit in cost– utility analysis is the quality-adjusted life year, but other measures include disability-adjusted life years (DALYs) and healthy year equivalents (HYEs).

References

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 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

Jill A. Hayden, Danielle A. van der Windt, Jennifer L. Cartwright, Pierre Côté, Claire Bombardier. (2013) Assessing Bias in Studies of Prognostic Factors. Ann Intern Med. 158: 280–286.

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

Fu R, Vandermeer BW, Shamliyan TA, et al. (2013) Handling Continuous Outcomes in Quantitative Synthesis In: Methods Guide for Effectiveness and Comparative Effectiveness Reviews [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2008-. (accessed 16th October 2025) http://www.ncbi.nlm.nih.gov/books/NBK154408/

Higgins JPT, Green S (editors) (2011) Cochrane Handbook for Systematic Reviews

of Interventions Version 6.5 [updated 2034] The Cochrane Collaboration. Available from https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current (accessed on 16th October 2025)

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

McGowan 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

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

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

Huan Wang, Yancong Chen, Yali Lin, Julius Abesig, Irene XY Wu, Wilson Tam (2021) The methodological quality of individual participant data meta-analysis on

intervention effects: systematic review. Available from https://pubmed.ncbi.nlm.nih.gov/33875446/ (accessed on 16th October 2025)