

Draft

# Early and locally advanced breast cancer: diagnosis and management

NICE guideline: methods

*NICE guideline NG101*

*Methods (lymphoedema update)*

*September 2024*

*Draft for Consultation*

*Evidence reviews were developed by  
Guideline development team*



## **Disclaimer**

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

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

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 All NICE guidance is subject to regular review and may be updated or withdrawn.

## **Copyright**

© NICE 2024. All rights reserved. Subject to Notice of rights.

ISBN:

# Contents

<b>Development of the guideline</b> .....	<b>5</b>
What this guideline covers .....	5
What this guideline does not cover .....	5
<b>Methods</b> .....	<b>6</b>
Developing the review questions and outcomes .....	6
Reviewing research evidence .....	6
Review protocols .....	6
Searching for evidence .....	6
Selecting studies for inclusion.....	6
Incorporating published evidence syntheses .....	7
Methods of combining evidence.....	8
Data synthesis for intervention studies .....	8
Appraising the quality of evidence .....	10
Intervention studies (relative effect estimates).....	10
Minimally important differences (MIDs) and clinical decision thresholds .....	10
GRADE for intervention studies analysed using pairwise analysis.....	12
Reviewing economic evidence .....	13
Inclusion and exclusion of economic studies .....	13
Appraising the quality of economic evidence.....	13

# 1 **Development of the guideline**

## 2 **What this guideline covers**

3 This guideline update will cover evidence and recommendations related specifically to  
4 preventing and managing lymphoedema in people with early, locally advanced, and  
5 advanced breast cancer. The update will focus on reviewing the evidence and updating or  
6 creating new recommendations in the lymphoedema section of the existing NICE guidelines  
7 on early and locally advanced breast cancer (NG101) and advanced breast cancer (CG81).

8 The update will look at the effectiveness and cost-effectiveness of non-pharmacological  
9 strategies for:

- 10 1. Reducing the risk of developing lymphoedema in people who have or have had  
11 breast cancer.
- 12 2. Managing lymphoedema in people who have or have had breast cancer.

## 13 **What this guideline does not cover**

14 This guideline update will not review evidence or update recommendations in other sections  
15 of NG101 (published in 2018) and CG81 (published in 2009) beyond the lymphoedema  
16 section. Existing recommendations in sections that are not being updated will be retained  
17 from the previously published versions of the guidelines.

## 1 **Methods**

2 This guideline was developed using the methods described in the [2024 NICE guidelines](#)  
3 [manual](#).

4 Declarations of interest were recorded according to the NICE conflicts of interest policy.  
5

## 6 **Developing the review questions and outcomes**

7 The 2 review questions developed for this guideline were based on the key areas identified  
8 in the guideline [scope](#). They were drafted by the NICE guideline development team and  
9 refined and validated by the guideline committee.

10 The review questions were based on the following frameworks:

- 11 • Population, Intervention, Comparator and Outcome and Study type (PICOS) for  
12 reviews of interventions

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

## 15 **Reviewing research evidence**

### 16 **Review protocols**

17 Review protocols were developed with the guideline committee to outline the inclusion and  
18 exclusion criteria used to select studies for each evidence review. The review protocols  
19 were prospectively registered in the [PROSPERO register of systematic reviews](#).  
20

21 PROSPERO Registration numbers:

- 22 • Review Protocol 1 – Prevention of lymphoedema: CRD42024521526
- 23 • Review Protocol 2 – Management of lymphoedema: CRD42024521515  
24

### 25 **Searching for evidence**

26 Evidence was searched for each review question using the methods specified in the [2024](#)  
27 [NICE guidelines manual](#).

### 28 **Selecting studies for inclusion**

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

35 The full text of potentially eligible studies was retrieved and assessed according to the  
36 criteria specified in the review protocol. A standardised form was used to extract data from

1 included studies. Study investigators were contacted for missing data when time and  
2 resources allowed (when this occurred, this was noted in the evidence review and relevant  
3 data was included).

#### 4 **Incorporating published evidence syntheses**

5 If published evidence syntheses were identified sufficiently early in the review process (for  
6 example, from the surveillance review or early in the database search), they were  
7 considered for use as the primary source of data, rather than extracting information from  
8 primary studies. Syntheses considered for inclusion in this way were quality assessed to  
9 assess their suitability using the appropriate checklist, as outlined in

10 Table 1. Note that this quality assessment was solely used to assess the quality of the  
11 synthesis in order to decide whether it could be used as a source of data, as outlined in  
12 Table 2, not the quality of evidence contained within it, which was assessed in the usual way  
13 as outlined in the section on 'Appraising the quality of evidence'.

14 **Table 1: Checklists for published evidence syntheses**

Type of synthesis	Checklist for quality appraisal
Systematic review of quantitative evidence	ROBIS
Network meta-analysis	Modified version of the PRISMA NMA tool (see appendix K of 'Developing NICE guidelines, the manual')

15

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

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

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

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

34 The way that a published evidence synthesis was used in the evidence review depended on  
35 its quality and applicability, as defined in Table 2. When published evidence syntheses were  
36 used as a source of primary data, data from these evidence syntheses were quality

1 assessed and presented in GRADE in the same way as if data had been extracted from  
 2 primary studies. In questions where data was extracted from both systematic reviews and  
 3 primary studies, these were checked to ensure none of the data had been double counted  
 4 through this process.

5 **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.

6

7

## 8 **Methods of combining evidence**

### 9 **Data synthesis for intervention studies**

10 Where possible, meta-analyses were conducted to combine the results of quantitative  
 11 studies for each outcome. When there were 2 treatment alternatives, pairwise meta-analysis  
 12 was used to compare interventions.

13



## 1 **Pairwise meta-analysis**

2 Pairwise meta-analyses were performed in Cochrane Review Manager V5.3. A pooled  
3 relative risk was calculated for dichotomous outcomes (using the Mantel–Haenszel method)  
4 reporting numbers of people having an event. Both relative and absolute risks were  
5 presented, with absolute risks calculated by applying the relative risk to the risk in the  
6 comparator arm of the meta-analysis (calculated as the total number events in the  
7 comparator arms of studies in the meta-analysis divided by the total number of participants  
8 in the comparator arms of studies in the meta-analysis).

9 A pooled mean difference was calculated for continuous outcomes (using the inverse  
10 variance method) when the same scale was used to measure an outcome across different  
11 studies. Where different studies presented continuous data measuring the same outcome  
12 but using different numerical scales (e.g. a 0-10 and a 0-100 visual analogue scale), these  
13 outcomes were all converted to the same scale before meta-analysis was conducted on the  
14 mean differences.

15 Where outcomes measured the same underlying construct but used different  
16 instruments/metrics, data were analysed using standardised mean differences (SMDs,  
17 Hedges' g), as implemented in Review manager version 5. Alternative approaches to  
18 standardisation described in the NICE technical support unit guideline methodology  
19 document on meta-analysis of continuous outcomes were used when this was needed for  
20 consistency with a network meta-analysis.

21 For continuous outcomes analysed as mean differences, change from baseline values were  
22 used in the meta-analysis if they were accompanied by a measure of spread (for example  
23 standard deviation). Where change from baseline (accompanied by a measure of spread)  
24 were not reported, the corresponding values at the timepoint of interest were used. If only a  
25 subset of trials reported change from baseline data, final timepoint values were combined  
26 with change from baseline values to produce summary estimates of effect. For continuous  
27 outcomes analysed as standardised mean differences this was not possible. In this case, if  
28 all studies reported final timepoint data, this was used in the analysis. If some studies only  
29 reported data as a change from baseline, analysis was done on these data, and for studies  
30 where only baseline and final time point values were available, change from baseline  
31 standard deviations were estimated, assuming a correlation coefficient derived from studies  
32 reporting both baseline and endpoint data, or if no such studies were available, assuming a  
33 correlation of 0.5 as a conservative estimate (Follman et al., 1992; Fu et al., 2013). In cases  
34 where SMDs were used they were back converted to a single scale to aid interpretation by  
35 the committee where possible, and when it was considered useful for decision making.

36 Random effects models were fitted when significant between-study heterogeneity in  
37 methodology, population, intervention or comparator was identified by the reviewer in  
38 advance of data analysis. This decision was made and recorded before any data analysis  
39 was undertaken. For all other syntheses, fixed- and random-effects models were fitted, with  
40 the presented analysis dependent on the degree of heterogeneity in the assembled  
41 evidence. Fixed-effects models were the preferred choice to report, but in situations where  
42 the assumption of a shared mean for fixed-effects model were clearly not met, even after  
43 appropriate pre-specified subgroup analyses were conducted, random-effects results are  
44 presented. Fixed-effects models were deemed to be inappropriate if there was significant  
45 statistical heterogeneity in the meta-analysis, defined as  $I^2 \geq 50\%$ .

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

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

## 7 **Appraising the quality of evidence**

### 8 **Intervention studies (relative effect estimates)**

9 RCTs and quasi-randomised controlled trials were quality assessed using the Cochrane Risk  
10 of Bias Tool 2. Non-randomised controlled trials and cohort studies were quality assessed  
11 using the ROBINS-I tool. Other study types (for example controlled before and after studies)  
12 were assessed using the preferred option specified in the [NICE guidelines manual 2024](#)  
13 [\(appendix H\)](#). Evidence on each outcome for each individual study was classified into one of  
14 the following groups:

- 15 • Low risk of bias – The true effect size for the study is likely to be close to the estimated  
16 effect size.
- 17 • Moderate risk of bias – There is a possibility the true effect size for the study is  
18 substantially different to the estimated effect size.
- 19 • High risk of bias – It is likely the true effect size for the study is substantially different to  
20 the estimated effect size.
- 21 • Critical risk of bias (ROBINS-I only) - It is very likely the true effect size for the study is  
22 substantially different to the estimated effect size.

23

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

28

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

### 35 **Minimally important differences (MIDs) and clinical decision thresholds**

36 The Core Outcome Measures in Effectiveness Trials (COMET) database was searched to  
37 identify published minimal clinically important difference thresholds relevant to this guideline  
38 that might aid the committee in identifying clinical decision thresholds for the purpose of  
39 GRADE. Identified MIDs were assessed to ensure they had been developed and validated in

1 a methodologically rigorous way, and were applicable to the populations, interventions and  
 2 outcomes specified in this guideline. In addition, the Guideline Committee were asked to  
 3 prospectively specify any outcomes where they felt a consensus clinical decision threshold  
 4 could be defined from their experience. Any questions looking to evaluate non-inferiority  
 5 (that one treatment is not meaningfully worse than another) required a clinical decision  
 6 threshold to be defined to act as a non-inferiority margin.

7 Clinical decision thresholds were used to assess imprecision using GRADE and aid  
 8 interpretation of the size of effects for different outcomes. Clinical decision threshold that  
 9 was used in the guideline are given in Table 3 and also reported in the relevant evidence  
 10 reviews.

11 **Table 3 Identified Clinical decision thresholds**

Outcome	Clinical decision threshold	Source
DASH scale	7-point difference: MD -7 to +7 points	Bruce J, Mazuquin B, Mistry P, Rees S, Canaway A, Hossain A, et al. Exercise to prevent shoulder problems after breast cancer surgery: the PROSPER RCT. <i>Health Technol Assess</i> 2022;26(15).
SF-36	one-half of an SD	Reeve BB, Potosky AL, Smith AW, Han PK, Hays RD, Davis WW, Arora NK, Haffer SC, Clauser SB. Impact of cancer on health-related quality of life of older Americans. <i>JNCI: Journal of the National Cancer Institute</i> . 2009 Jun 16;101(12):860-8.
QuickDASH scale	8-point difference: MD -8 to +8 points	Mintken PE, Glynn P, Cleland JA. Psychometric properties of the shortened disabilities of the Arm, Shoulder, and Hand Questionnaire (QuickDASH) and Numeric Pain Rating Scale in patients with shoulder pain. <i>J Shoulder Elbow Surg</i> . 2009 Nov-Dec;18(6):920-6.
Pain 100-mm VAS	More or less pain: 13-mm change in score	Gallagher EJ, Liebman M, Bijur PE. Prospective validation of clinically important changes in pain severity measured on a visual analog scale. <i>Ann Emerg Med</i> . 2001 Dec;38(6):633-8.
FACT-B total	7-8 points	Evidence review A: surgery to the breast (NG101 2018 update)
EORTC-QLQ-C30	-8 to 12 for global quality of life	Musoro JZ, Coens C, Fiteni F, Katarzyna P, Cardoso F, Russell NS, King MT, Cocks K, Sprangers MA, Groenvold M, Velikova G, Flechtner HH, Bottomley A; EORTC Breast and Quality of Life Groups. Minimally Important Differences for Interpreting EORTC QLQ-C30 Scores in Patients With Advanced Breast Cancer. <i>JNCI Cancer Spectr</i> . 2019 Jun 4;3(3):pkz037.

12 For continuous outcomes expressed as a mean difference where no other clinical  
 13 decision threshold was available, where a mean difference (MD) was reported, the NICE  
 14 default clinical decision threshold of 0.5 of the standard deviation (SD) of the control  
 15 group for each outcome was used. Where the SD was not reported, the line of no effect  
 16 was used as a clinical decision threshold and a sample size of n <400 was  
 17 used to provide the second domain to downgrade for imprecision. Where a standardised

1 mean difference (SMD) was reported, the NICE default of +/-0.5 was used for the clinical  
 2 decision thresholds. (Norman et al. 2003. For SMDs that were back converted to one of  
 3 the original scales to aid interpretation, rating of imprecision was carried out before back  
 4 calculation. For relative risks and hazard ratios, where no other clinical decision  
 5 threshold was available, a default clinical decision threshold for dichotomous outcomes  
 6 of 0.8 to 1.25 was used. Odds ratios were converted to risk ratios before presentation to  
 7 the committee to aid interpretation.

## 8 GRADE for intervention studies analysed using pairwise analysis

9 GRADE was used to assess the quality of evidence for the outcomes specified in the review  
 10 protocol. Data from randomised controlled trials, non-randomised controlled trials and cohort  
 11 studies (which were quality assessed using the Cochrane risk of bias tool or ROBINS-I) were  
 12 initially rated as high quality while data from other study types were initially rated as low  
 13 quality. The quality of the evidence for each outcome was downgraded or not from this initial  
 14 point, based on the criteria given in [Table 4](#). These criteria were used to apply preliminary  
 15 ratings, but were overridden in cases where, in the view of the analyst or committee the  
 16 uncertainty identified was unlikely to have a meaningful impact on decision making.

17 The following criteria were used to interpret the effect (column of 'Interpretation of effect') in  
 18 the summary GRADE tables: For all outcomes, evidence statements are divided into 2  
 19 groups as follows:

- 20 • We state that the evidence showed that there is an effect if the 95% CI does not  
 21 cross the line of no effect
- 22 • The evidence could not differentiate between comparators if the 95% CI crosses the  
 23 line of no effect

24

25 **Table 4: Rationale for downgrading quality of evidence for intervention studies**

GRADE criteria	Reasons for downgrading quality
Risk of bias	<p>Not serious: If less than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the overall outcome was not downgraded.</p> <p>Serious: If greater than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the outcome was downgraded one level.</p> <p>Very serious: If greater than 33.3% of the weight in a meta-analysis came from studies at high risk of bias, the outcome was downgraded two levels.</p> <p>Extremely serious: If greater than 33.3% of the weight in a meta-analysis came from studies at critical risk of bias, the outcome was downgraded three levels</p>
Indirectness	<p>Not serious: If less than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the overall outcome was not downgraded.</p> <p>Serious: If greater than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the outcome was downgraded one level.</p> <p>Very serious: If greater than 33.3% of the weight in a meta-analysis came from indirect studies, the outcome was downgraded two levels.</p>
Inconsistency	Concerns about inconsistency of effects across studies, occurring when there is unexplained variability in the treatment effect demonstrated across studies

GRADE criteria	Reasons for downgrading quality
	<p>(heterogeneity), after appropriate pre-specified subgroup analyses have been conducted. This was assessed using the <math>I^2</math> statistic.</p> <p>Not serious: If the <math>I^2</math> was less than 33.3%, the outcome was not downgraded.</p> <p>Serious: If the <math>I^2</math> was between 33.3% and 66.7%, the outcome was downgraded one level OR If data on the outcome was only available from one study.</p> <p>Very serious: If the <math>I^2</math> was greater than 66.7%, the outcome was downgraded two levels.</p>
Imprecision	<p>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.</p> <p>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.</p> <p>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.</p>
Publication bias	<p>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.</p>

## 1 Reviewing economic evidence

### 2 Inclusion and exclusion of economic studies

3 Literature reviews seeking to identify published cost–utility analyses of relevance to the  
4 issues under consideration were conducted for all questions. In each case, the search  
5 undertaken for the clinical review was modified, retaining population and intervention  
6 descriptors, but removing any study-design filter and adding a filter designed to identify  
7 relevant health economic analyses. In assessing studies for inclusion, population,  
8 intervention and comparator, criteria were always identical to those used in the parallel  
9 clinical search; only cost–utility analyses were included. Economic evidence profiles,  
10 including critical appraisal according to the Guidelines manual, were completed for included  
11 studies.

### 12 Appraising the quality of economic evidence

13 Economic studies identified through a systematic search of the literature were appraised  
14 using a methodology checklist designed for economic evaluations (NICE guidelines manual;  
15 2014). This checklist is not intended to judge the quality of a study per se, but to determine

1 whether an existing economic evaluation is useful to inform the decision-making of the  
2 committee for a specific topic within the guideline.

3 There are 2 parts of the appraisal process. The first step is to assess applicability (that is,  
4 the relevance of the study to the specific guideline topic and the NICE reference case);  
5 evaluations are categorised according to the criteria in Table 5.

#### 6 **Table 5 Applicability criteria**

Level	Explanation
Directly applicable	The study meets all applicability criteria, or fails to meet one or more applicability criteria but this is unlikely to change the conclusions about cost effectiveness
Partially applicable	The study fails to meet one or more applicability criteria, and this could change the conclusions about cost effectiveness
Not applicable	The study fails to meet one or more applicability criteria, and this is likely to change the conclusions about cost effectiveness. These studies are excluded from further consideration

7 In the second step, only those studies deemed directly or partially applicable are further  
8 assessed for limitations (that is, methodological quality); see categorisation criteria in Table  
9 6.

#### 10 **Table 6 Methodological criteria**

Level	Explanation
Minor limitations	Meets all quality criteria, or fails to meet one or more quality criteria but this is unlikely to change the conclusions about cost effectiveness
Potentially serious limitations	Fails to meet one or more quality criteria and this could change the conclusions about cost effectiveness
Very serious limitations	Fails to meet one or more quality criteria and this is highly likely to change the conclusions about cost effectiveness. Such studies should usually be excluded from further consideration

11 Where relevant, a summary of the main findings from the systematic search, review and  
12 appraisal of economic evidence is presented in an economic evidence profile alongside the  
13 clinical evidence.  
14

## 1 **References**

2

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

5

6 Fu R, Vandermeer BW, Shamliyan TA, et al. (2013) Handling Continuous Outcomes in  
7 Quantitative Synthesis In: *Methods Guide for Effectiveness and Comparative Effectiveness*  
8 *Reviews* [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2008-  
9 . Available from: <http://www.ncbi.nlm.nih.gov/books/NBK154408/>

10

11 Norman G., Sloan JA., Wyrwich KW. (2003) Interpretation of changes in health-related  
12 quality of life: the remarkable universality of half a standard deviation. *Med Care* 41(5):582-  
13 92.

14