**Supplementary Materials:** *Reviewer Guidelines for Full Paper Screening, Data Extraction, and Quality Assessment*

This document contains the instructions to complete the full paper screening, data extraction, and quality assessments for a PROSPERO-registered (CRD42023484924) systematic review and meta-analysis. These steps immediately followed the title and abstract screening and were applied to all full papers from academic and grey literature searches. The results from completing this process are reported in the published manuscript, Appendix, and other Supplementary Materials.

The main reviewer screened the full papers for inclusion. They also completed data extraction and quality assessments for the full set of included studies. This process was completed in Google Sheets and is shared on ORDA as an Excel file. A second reviewer completed the data extraction and quality assessment processes for approximately 33% of the included papers. The included papers were organised alphabetically, and then a sample for the second reviewer was randomly selected using an online random number generator: https://www.calculator.net/random-number-generator.html The second reviewer had access to a separate Google Drive folder containing a copy of this guideline document, a Google Sheet document with a blank copy of the data extraction form and CASP tables, blank CASP forms, and copies of the randomly selected papers to complete the data extraction and quality assessments independently.

**Screening Full Papers for Inclusion**

In the data extraction table, first, check any notes from the title/abstract screening column. These notes are intended to guide the reviewer to specific sections of the paper when screening. Then use the following questions as a quick guide for inclusion. If the answer to any of these six questions is no, exclude the paper at this stage.

1. Are the papers fully in English?
2. Are the participants human?
3. Are there participants with a diagnosis of anorexia nervosa?
4. Are the participants also outpatients, as per the inclusion criteria definition (Table E.1)?
5. Is the intervention CBT as per the inclusion criteria definition (Table E.1)?
6. Is the study primary research which reports pre-and-post-intervention outcomes for at least one key variable of either weight or disordered eating symptoms (Table E.1)?

**Excluded Papers**

For any excluded papers, in the data extraction sheet:

* Mark N in the Include study? column.
* Add the reasons for exclusion in the next column, following the guidelines in Table E.2.
* Mark the whole row in red.

**Included Papers**

If the paper is not excluded at this stage, the reviewer needs to read the full paper carefully and check it against the inclusion and exclusion criteria in Table E.1 as they complete the data extraction process. If the paper is to be included:

* Complete the full data extraction process as outlined in Table E.2.
* Mark Y in the Include Study? column.
* Mark the whole row in green.
* Conduct the quality assessment process.

If, at any time, reading highlights the need to exclude a paper, it is not necessary to continue with the data extraction process. The above steps for excluding a paper can be followed.

**Table E.1**

# *Inclusion and exclusion criteria*

|  |  |  |
| --- | --- | --- |
| **PICO** | **Inclusion Criteria** | **Exclusion Criteria** |
| *Population* | * Humans. * Primary clinical diagnosis of anorexia nervosa (any subtype). * Subclinical patients included in analyses if they met the key criterion of being underweight and had all but one other diagnostic criterion for anorexia nervosa. * Outpatients. * Any other characteristics (e.g., age, sex, ethnicity). | * Non-human animals. * Known comorbid diagnoses that affect CBT delivery (i.e., psychosis, learning disabilities). * Other primary diagnoses of eating disorders like atypical anorexia nervosa, bulimia nervosa, binge-eating disorder, ARFID, EDNOS, or OSFED). * In-patients or day-patients. * Outpatient therapy immediately following in-patient therapy. * Studies where it is unclear whether patients are outpatients. |
| *Intervention* | * Primary studies. * First- or second-wave CBT (e.g., Enhanced CBT (CBT-E), Cognitive Therapies (CT), Behaviour or Behavioural Therapies (BT), Exposure Therapies, Body Image Therapies). | * Non-CBT interventions (e.g., Family Therapy, Psychotherapy). * Intensive CBT incorporating both in-patient and day-patient treatment. * Third-wave CBT (e.g., Acceptance and Commitment Therapy (ACT), Cognitive Analytical Therapy (CAT), Cognitive Remediation Therapy (CRT), Mindfulness Therapies). |
| *Comparison* | * Original primary research studies. * Any quantitative design. * Sample >1 to allow for calculation of effect size. * Outcomes reported separately by diagnosis. * Published or non-published papers and reports. * With and without control or comparator groups. * Any year of publication. * Any country or original language, but the full paper must be accessible in English. | * Non-primary research studies. * Qualitative studies. * Case studies, including single-case designs. * Studies that do not report outcomes by diagnosis. * Non-English language publications. |
| *Outcomes* | * Reports within-subject pre-to-post-treatment scores for at least one primary outcome measure of either weight or eating disorder symptoms. | * Does not report quantifiable outcomes in at least one primary outcome measure. |

*Note:* PICO criteria refer to the Population, Intervention, Comparison, Outcomes criteria reported by Eriksen & Frandsen (2018) and Higgins et al. (2022).

# **Completing the Data Extraction Table**

The following approach will help the reviewer screen the full papers and extract data as quickly and efficiently as possible. Full details of the required data for each of the columns in the data extraction table are given in Table E.2. If, at any time, a lack of appropriate data suggests the need to exclude a paper, it is not necessary to continue the data extraction process, and the previous exclusion guidelines can be followed.

**Suggested Extraction Process**

1. Record the pre-post primary outcomes for at least one key outcome variable, following the guidelines in Table E.2 to collect sufficient data for effect size calculations. *Remember, this outcome data must be only for patients with anorexia nervosa in receipt of outpatient CBT. If such outcomes are not available, exclude the study.*
2. Record sample sizes appropriate to the pre-post primary outcomes. *For example, if the outcome data is for completers only, then make sure to note the completer sample size alongside other sample size data.*
3. Check any outstanding notes from previous stages of reviewing. If there are any doubts about inclusion from the earlier title/abstract screening notes, then address the specific data that will give you an answer.
4. Complete all other columns in the data extraction form with as much data as possible.
5. For all fields where there is missing data, mark it as NA or unknown, as appropriate.

# **Table E.2**

# *Coding protocol for data extraction*

| **Column Name** | **Instructions for Completing Data Extraction** |
| --- | --- |
| Study Author/s | Authors’ names written in citation format according to APA 7th Edition formatting. |
| Publication Date | Year of publication (YYYY). |
| Article Title | Title of the study/article/report. |
| DOI | DOI or web link, if available. |
| Publication Type | * Book Chapter * Conference Paper/Abstract * Journal Article * Organisational Report (e.g., governmental or charity report) * Magazine Article * Pre-Print * Thesis/Dissertation * Other (give details in the ‘Notes’ column). |
| Published Study | * Y if published * N if unpublished * Unknown |
| Study Location | Enter the name of the country or countries where the study was conducted (e.g., USA, UK, Italy, Japan, etc.). Record the names of multiple countries if the study was conducted in more than one location (e.g., Italy, UK). If the country of study is not indicated, enter the country of the first author (usually indicated by their institution) as a proxy for the likely study location. If this information is unavailable, mark it as unknown. |
| Study Design | * Case Series (i.e., studies with no control groups and non-random allocation often conducted in routine clinical practice). * Case Control (i.e., studies with a control or comparison group (e.g., other treatments, eating disorder diagnoses, or healthy controls) and either random or non-random allocation, often conducted in routine clinical practice). * Cohort (i.e., studies with no control groups and no random allocation, but may include a comparison group, often conducted in routine clinical practice from recruitment of consecutive referrals to services). * Randomised Controlled Trial (RCT) (i.e., studies to test the effectiveness of one or more interventions with comparison or control groups and random allocation). * Non-Randomised Controlled Trial (i.e., studies to test the effectiveness of one or more interventions with comparison or control groups but with non-random allocation). * Experiment (i.e., studies with specific experimental aims, usually with control groups and random allocation). * Other (Give details in the ‘Notes’ column).   Study design categories taken from Barker et al. (2016). |
| Participant Allocation | * Random (e.g., in an RCT or case-control study). * Non-random (e.g., only one group being studied, naturalistic observation studies conducted in clinical practice, consecutive clinical referrals in cohort studies). * Unknown (i.e., it is unclear whether participants were randomly allocated to groups). |
| Single or Multi-Centre Study | * Single Centre * Multi Centre * Unknown |
| Subsample Descriptor | This field is included to describe the specific subsample of participants to which the data recorded on this row relates. If there is more than one relevant subsample of participants in a study, add another row at this point.  When to add in additional rows for a study:   * Both completer and intention to treat (ITT) data are provided. * There is more than one group receiving CBT or receiving CBT in combination with some other treatment (e.g., two groups receiving CBT plus different types of medication). * Data is reported separately for patients in different countries.   In this field, clearly state the following information:   * The name of the subsample (e.g., CBT plus medication, or completers), * Whether the quantitative outcome data is calculated for completers or ITT (sometimes this is obvious as this is the subgrouping, but sometimes, as with the CBT plus medication example, it is not obvious. Check when entering the outcome data. * The sample size which applies to this subgroup (e.g., *N* = 25, completers only). |
| ***N* for Effect Size Calculations** In this section, record the reported values for as applicable to the subsample recorded on each row. | |
| Total Number of Participants in the Whole Study Sample *(irrespective of diagnosis)* | Total number of participants in the whole study, regardless of diagnosis or whether they are experimental or control groups. |
| Total Number of Participants with Anorexia Nervosa Allocated to CBT | Total number of participants with anorexia nervosa in the study who were allocated to receive CBT. This *N* represents the ITT sample (i.e., those participants who started the study but did not necessarily complete it). |
| Total Number of Participants with Anorexia Nervosa Restrictive Type Allocated to Receive CBT | Subsample size of any participants with anorexia nervosa restrictive type (AN-R) allocated to receive CBT. |
| Total Number of Participants with Anorexia Nervosa Binge-Purge Subtype Allocated to CBT | Subsample size of any participants with anorexia nervosa restrictive type (AN-BP, AN-P) allocated to receive CBT. |
| Total Number of Participants with Anorexia Nervosa Allocated to CBT and Lost to Attrition | Total number of participants with anorexia nervosa allocated to CBT and lost to attrition. If not reported, this can be calculating by taking the total number of participants with anorexia nervosa completing CBT from the total number of participants with anorexia nervosa allocated to CBT and lost to attrition. |
| Total Number of Participants with Anorexia Nervosa Completing CBT | Total number of participants with anorexia nervosa who have completed a full course of CBT in the study. This represents the *N* for completer data (i.e., patients who have completed a full course of CBT). This can be calculated by taking the total number of participants with anorexia nervosa allocated to CBT and lost to attrition from the total number of participants with anorexia nervosa completing CBT. |
| **Patient Characteristics**  These figures are commonly stated for the whole sample and not necessarily by diagnosis. For this review, only record characteristics for the appropriate subsample of patients with anorexia nervosa on each row. Otherwise, mark as unknown/NA. If a study does not report patient characteristic information but reports key outcome data, it can still be included in the review. | |
| Mean Age | Mean age (in years). |
| Sex | Percentage of female participants. This can be calculated from the number of female participants divided by the total number of participants and multiplied by 100%. |
| Mean Baseline BMI | Mean baseline BMI, regardless of subtype. If BMI is not reported, record alternative measures of weight (e.g., mean weight in kilograms and height in metres, adjusted weight for height). Indicate clearly in brackets any alternative measure being recorded. |
| Mean Duration of Illness | Mean duration of illness. Indicate in brackets whether in years or months. |
| Ethnicity | Percentage of participants for whom their ethnicity is listed as white. |
| Ethnicities Reported | List any ethnicities reported using the categories below with a comma between each (e.g., Black, Asian, White).   * Black (e.g., Caribbean, African, Black British/American, or any other Black background). * Asian (e.g., Indian, Pakistani, Bangladeshi, Chinese, Asian British/American, or any other Asian background). * Arabic (e.g., Middle East). * White (e.g., Any Caucasian ethnicity (e.g., British, American), including Gypsy/Irish Traveller, Roma, or any other White background). * Mixed or Multiple Ethnicity (e.g., White and Black Caribbean. White and Black African, White and Asian, or any other Mixed or Multiple Ethnic backgrounds). * Other (give details in brackets).   Categories adapted from GOV.UK (2021). |
| Autism Spectrum Disorder Diagnoses Reported | * Y if reported * N if not reported |
| %age of Anorexia Nervosa CBT Participants with Autism Spectrum Disorder | For those where a subsample does have ASD, note the percentage of anorexia nervosa patients receiving CBT who have autism spectrum disorder. If ASD is not reported, state NA. |
| Name of Therapy | Name used to describe the CBT therapy in the study. Examples include:   * CBT for Eating Disorders (CBT-ED) * Enhanced CBT (CBT-E) * Behavioural Therapy (BT) * Exposure with Response Prevention (ERP)   Enter the name exactly as it is written in the paper and include both the full therapy name and any acronym used in brackets as above. |
| Individual or Group | * Individual * Group   Unless otherwise stated, mark the therapy as individual as this appears to be the standard mode of delivery in most studies, though this information is not always explicitly stated. |
| In-Person, Online, or Telephone | * In person * Online * Telephone   Unless otherwise stated, mark therapy is in person as this appears to be the standard mode of delivery in most studies, though this information is not always explicitly stated. |
| Total Number of Sessions | Total number of sessions of CBT either planned or received. Give as much additional information as possible in brackets (e.g., mean, mode, or median, protocol or received number of sessions). |
| Frequency of Sessions | * More than Twice Weekly * Twice Weekly * Weekly * Fortnightly * Monthly * Less than Monthly * Unknown |
| Manual Used | * Y if a manual was used or protocol followed. * N if a manual was not stated as being used or protocol was not clearly followed.   Make any additional notes in the Included Interventions box, along with any adaptations made for the study if necessary. |
| Name of Manual Authors and Date of Publication | Name of manual authors and the date of its publication in brackets in APA 7th edition citation style, if reported. |
| Any Amendments to Manual Protocol | Note any amendments to the manual protocol as reported in the paper, if reported. |
| Included Interventions | For non-manual-based studies only, list any included interventions as written in the paper. This will help to generate a picture of the CBT being used when there is no manual to refer to. |
| Order of Contents: Behavioural Interventions Ahead of Cognitive Interventions | * Y if behavioural interventions were conducted before the cognitive interventions. * N if cognitive interventions were conducted before behavioural interventions. * Mark as Manual if this order is unclear in the paper but a manual was used. The main reviewer will add this detail for each study that is manual based after reading the relevant manuals. * NA if no manual was used. |
| Therapist Experience | * Experienced (i.e., qualified professionals like doctors, nursing staff, or clinical psychologists). * Novice (i.e., individuals delivering therapy without formal clinical qualifications, but likely with some psychology background or equivalent, such as PhD Researchers or Research Assistants). * Both (for studies with both experienced and novice therapists). * Unknown (if unclear/unreported). |
| Key Findings | Indicate in words what the study’s key findings/outcomes were. These are usually highlighted in the abstract and the first paragraph of the discussion. Wherever possible, note the key findings relevant to this review’s primary (i.e., BMI/weight and eating disorder symptoms) and secondary (i.e., depression, anxiety, and quality of life) outcomes of interest. |
| **Data for Effect Size Information** Please record one effect size (or one measure’s set of data to calculate effect size) for each construct of interest (i.e., weight, eating disorder symptoms, depression, anxiety, or quality of life). This means if there are multiple measures of the same construct, you must choose the most appropriate based on the instructions below listed in order (i.e., 1 as top priority). These measures have been chosen as the first option because of their popularity in the eating disorder literature.  For each of the following measures, record only global scores. If the scores given are not global scores but subscale scores, please do not record them. This is because global and subscale scores are not comparable, and trying to calculate global from subscale scores may result in errors being made that then affect the meta-analyses.  In cases where there are two or more different study conditions, each of which contains CBT, please enter the scores for each on separate lines (i.e., one line for each condition) as mentioned earlier, completing the Subsample Descriptor column clearly so that it is clear for whom this data applies.  **Guidelines for Recording Appropriate Measures when there are Multiple Measures**  Weight   1. If reported, record scores from the Body Mass Index (BMI)measure. 2. If not, record the measure that the authors define in the text as their primary outcome variable for this construct. 3. If not, record the measure that the authors use in their main tables for this construct. 4. If there is no measure recorded for this construct, mark NA for this and subsequent related columns.   Eating Disorder Symptoms   1. If reported, record global/total scores from the Eating Disorder Examination Questionnaire (EDE-Q) (Mond et al., 2004) measure. 2. If not, record global/total scores for the measure that the authors define in the text as their primary outcome variable for this construct. 3. If not, record global/total scores for the measure that the authors use in their main tables for this construct. 4. If there is no measure recorded for this construct, mark NA for this and subsequent related columns.   Depression   1. If reported, record global/total scores from the PHQ-9 (Patient Health Questionnaire-9)(Kroenke et al., 2001)measure. 2. If not, record global/total scores for the measure that the authors define in the text as their primary outcome variable for this construct. 3. If not, record global/total scores for the measure that the authors use in their main tables for this construct. 4. If there is no measure recorded for this construct, mark NA for this and subsequent related columns.   Anxiety   1. If reported, record global/total scores from the Generalised Anxiety Disorder Questionnaire (GAD-7)(Spitzer et al., 2006) measure. 2. If not, record global/total scores for the measure that the authors define in the text as their primary outcome variable for this construct. 3. If not, record global/total scores for the measure that the authors use in their main tables for this construct. 4. If there is no measure recorded for this construct, mark NA for this and subsequent related columns.   Quality of Life   1. If reported, record global/total scores from the Clinical Impairment Assessment (CIA)(Bohn et al., 2008) measure. 2. If not, record global/total scores for the measure that the authors define in the text as their primary outcome variable for this construct. 3. If not, record global/total scores for the measure that the authors use in their main tables for this construct. 4. If there is no measure recorded for this construct, mark NA for this and subsequent related columns.   For each of the outcome measures, record the relevant information for the subsample on that row according to the below guidelines. Recording the pre-and-post-treatment means and standard deviations are the minimum data requirements for effect size calculations. Additional data should be recorded in as much detail as possible. | |
| Outcome Reported | * Y if a relevant outcome measure is reported for this construct (use the above guidelines to select the most appropriate measure). * N if there is no relevant outcome reported for this construct.   If there is no relevant measure for this construct, write NA in all following boxes and then move on to the next construct. |
| Name of Outcome Measure | Record the name of the outcome measure being recorded for this construct and any abbreviation used (e.g., Eating Disorder Examination Questionnaire (EDE-Q)) exactly as it is recorded in the paper. |
| Mean Pre-Intervention Outcome Score | Record the mean pre-intervention outcome score.  Note: Not all studies will necessarily report this data. If non-parametric alternatives are reported, add them to this column and make a note of what measure it is in brackets (e.g., median). |
| SD Pre-Intervention Outcome Score | Record the standard deviation (SD) for the pre-intervention outcome score.   * Not all studies will necessarily report this data. If non-parametric alternatives are reported, add them to this column and make a note of what measure it is in brackets (e.g., quartiles or inter-quartile range (IQR)). * Record the standard error (SE) if there is no SD reported. The standard deviation can be estimated using the SE and sample size (N): SD = SE x square root of N. This estimate assumes approximate normal distribution of the data and gives an estimate SD for the sample’s individual data points*.* |
| Mean Post-Intervention Outcome Score | Report the mean post-intervention outcome score. |
| SD Post-Intervention Outcome Score | Report the standard deviation for the post-intervention outcome score. |
| Backup Information: Alternative Scores for Effect Size Calculations | In some studies, means and standard deviations may not be provided. In this case, record any statistical scores which may be used as an alternative to calculate effect sizes. In Meta-Essentials, the two main statistics that can be used are F-values and t-values. Note which statistic the scores represent in brackets (e.g., 0.6 (t-value)). |
| Pre-Post Correlation Score | Note the correlation coefficient between pre-and-post-intervention outcome scores if reported.  This information is crucial for meta-analysis as it is needed to calculate effect sizes. However, this statistic is commonly not reported explicitly. An empirically derived correlation coefficient estimate of *r* = 0.57 (Balk et al., 2012; Gaskell et al., 2023) will be used to calculate an effect size for the included studies if not reported to allow for replication. |
| Type of Pre-Post Correlation Measure | If reported, note the type of correlation measure used. Examples include:   * Pearson’s correlation coefficient (r) * Spearman’s rank correlation coefficient (p or rs) * Point-biserial correlation coefficient (rpb) * Phi coefficient (ϕ) * Intraclass correlation coefficient (ICC) |
| Pre-Post Effect Size Score | If reported, note the effect size score. If an effect size is not reported, which is not typically the case, make sure that you have recorded the following information for an effect size to be estimated:   * Pre-and-post-intervention means and standard deviations. * Pre-and-post-intervention correlation coefficient (or estimate thereof). * The appropriate *N* for the specific subsample to whom the statistics above apply. |
| Type of Pre-Post Effect Size Measure | * Raw mean differences (D) * Cohen’s *d* (standardised mean differences) * Hedges’ *g* (standardised mean differences) * Cohen’s *f* (ANOVA-based ratio measure of effect size) * d-prime (standardised mean differences) * Response ratios (R) * Correlation coefficient (r) * Odds ratios (OR) * Risk ratios (RR) * Risk differences (RD) * Partial eta squared (η2p) * Other (Give details in the ‘Notes’ box) |
| Early Change Columns for Outcome Measures | Follow the above guidance to complete the relevant early change columns if scores are reported for early change outcomes. Early change scores are defined as any scores recorded between pre-and-post-intervention measures, typically in the first half of therapy. If there is more than one early change score, note the earliest early change indication scores in the following boxes. As per the pre-and-post-treatment outcome data, the most important data to record are the means and standard deviations (or equivalent). Record the effect sizes and correlation coefficients, if reported. The main reviewer will later estimate the effect sizes from this data. There is an additional ‘Timing of Early Change Score’ column. Record the timing of the early change score in this box (e.g., Session 6- or 4-weeks post-treatment start). |
| Follow-Up Columns for Outcomes Measures | Follow the above guidance to complete the relevant follow-up columns if scores are reported for follow-up outcomes. Follow-up scores are defined as any scores recorded after the post-intervention scores (i.e. after a full course of treatment is complete). If there is more than one follow-up score, note the last follow-up scores recorded (e.g., if there are six-month and one-year scores, note the scores recorded at one-year). As per the pre-and-post-treatment outcome data, the most important data to record are the means and standard deviations (or equivalent). Record the effect sizes and correlation coefficients, if reported. The main reviewer will later estimate the effect sizes from this data. There is an additional ‘Timing of Follow-Up Score’ column. Record the timing of the follow-up score in this box (e.g., one-year). |
| Notes from Title/Abstract Screening | Notes from the main reviewer’s initial title and abstract screening will be copied into this column to be used in the full paper screening process. |
| Include Study | * Y if studies are to be included * N if they are to be excluded * TBC if it is unclear whether to include the study (give details in the notes box) |
| Reasons for Exclusion | Give detailed reasons to be given underneath the brief reason taken from the below list.   * Not CBT * Not anorexia nervosa * Not a primary intervention-based research study (e.g., secondary data analysis) * Not outpatients (inpatients/day-patients, some excluded combination, or unclear whether outpatients or not) * Full paper not in English * Known comorbid diagnosis (that might affect how CBT is delivered) * Key outcomes not reported * Outcomes not reported by diagnosis * Outcomes not reported for only the CBT group * Outcomes not reported for outpatients only * Duplicate entry * Single case design * Outpatient therapy immediately following in-patient therapy * Other (give further details in brackets) |
| Other Notes | This section is for any other notes, whether those prompted by another question or any other pertinent information to the review.  Examples of notable details include:   * Decisions made during the data extraction process * Whether the correlation coefficient estimate is to be used * Notable patient characteristics or features of the intervention * Whether or not drugs or other therapies are being used alongside CBT |
| Type of Quality Assessment (CASP) Form Used | * CASP Cohort Studies * CASP Case-Control Studies * CASP RCTs |
| Quality Rating | * High * Medium * Low |

# **Conducting the Quality Assessment**

The following instructions are for completing the quality assessment forms for each of the included studies. You only need to complete a Critical Appraisal Skills Programme (CASP) form (Long et al., 2020), the associated CASP tables, and columns in the data extraction sheet when you are sure the study will be included.

*Note.* The CASP forms are downloadable directly from the website: <https://casp-uk.net/>. The additional tables are to be found in the Appendix and Supplementary Materials shared with the published paper.

**Instructions**

1. Choose the most appropriate CASP form for the study from the Google Drive templates folder based on the study design recorded during data extraction:

* CASP Case-Control Studies forms are to be used with case-control studies.
* CASP RCT forms are to be used with RCTs and non-RCTs only, bearing in mind that scores will be inherently lower for studies which are non-RCTs.
* CASP Cohort Studies forms are to be used with all other study designs. Most intervention studies are likely to fall into this category.

1. Make a copy of the CASP form template and save it as ‘STUDY AUTHORS\_DATE\_CASP FORM TYPE\_REVIEWER INITIALS’ (e.g., Smith et al\_2023\_CASP Cohort Studies Form\_HD) in the reviewer folder.
2. Write the name of the paper at the top of the CASP form.
3. Complete the questions in the CASP form. Most questions are self-explanatory and offer examples of what to look for in the form itself. More detailed guidelines for each specific form are given on the CASP website: https://casp-uk.net/how-to-use-checklist/ Question 12 of the CASP Cohort Studies Form is not phrased in a way that fits the choice of answers. For this review, this question should be interpreted as follows: *Are the implications of this study for practice identifiable?*
4. Record the CASP form results in the CASP table Google Sheet as follows:

* Find the CASP form table that correctly corresponds to the CASP form type.
* Add the author’s name/s and date of the paper in APA citation style to the Author/s column.
* Record the answers to the quantitative questions as indicated in each of the columns, typing Y for yes, N for no, and CT for any answers where it is not possible to tell from the information whether the paper contains the relevant information.

1. Add up the number of questions which were answered as yes.
2. Write this number in the total score column of the CASP table.

Use this number to calculate the percentage score as the total of yes responses divided by the total possible quantitative responses multiplied by 100%. Note that the CASP forms for RCTs are scored out of 13, case-control studies out of 10, and cohort studies out of 12.

1. Add the percentage to the column in the CASP table.
2. Rate the study as high, medium, or low quality and note this in the Quality Rating column of the CASP table as follows:

* High: ≥66.7%
* Medium: 33.4% - 66.6%
* Low: ≤33.3%

1. Transfer this information to the top of the CASP form so that there is a record of the outcomes on the form itself.
2. Record the Type of Quality Assessment (CASP) Form Used and the Quality Rating in the relevant columns of the data extraction sheet.

While the CASP checklists are recognised and used in clinical practice to evaluate the quality of intervention studies (Long et al., 2020), a key limitation is that these forms were not intended to be rated quantitatively. However, for the purpose of this review, a quantitative quality score will be calculated to give an approximate indication of quality. While this is not an accurate estimate of overall quality, this is a comparable estimate of quality across the included studies.

**References**

Balk, E.M., Earley, A., Patel, K., Trikalinos, T.A., & Dahabreh, I.J. (2012). *Empirical assessment of within-arm correlation imputation in trials of continuous outcomes.* Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services. https://www.ncbi.nlm.nih.gov/books/NBK115797/

Barker, C., Pistrang, N., & Elliott, R. (2016). *Research methods in clinical psychology: an introduction for students and practitioners* (3rd ed.). Wiley-Blackwell.

Bohn, K., Doll, H. A., Cooper, Z., O'Connor, M., Palmer, R. L., & Fairburn, C. G. (2008). The measurement of impairment due to eating disorder psychopathology. *Behaviour Research and Therapy*, *46*(10), 1105–1110. https://doi.org/10.1016/j.brat.2008.06.012

​​Eriksen, M. B., & Frandsen, T. F. (2018). The impact of patient, intervention, comparison, outcome (PICO) as a search strategy tool on literature search quality: a systematic review. *Journal of the Medical Library Association, 106*(4), 420–431. https://doi.org/10.5195/jmla.2018.345

GOV.UK (2021). *List of ethnic groups.* https://www.ethnicity-facts-figures.service.gov.uk/style-guide/ethnic-groups/

Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., Welch, V.A. (Eds), *Cochrane Handbook for Systematic Reviews of Interventions* (Version 6.4). https://www.training.cochrane.org/handbook

Gaskell, C., Simmonds-Buckley, M., Kellett, S., Stockton, C., Somerville, E., Rogerson, E., & Delgadillo, J. (2023). The effectiveness of psychological interventions delivered in routine practice: systematic review and meta-analysis. *Administration and Policy in Mental Health and Mental Services Research, 50*, 43-57. https://doi.org/10.1007/s10488-022-01225-y

Kroenke, K., Spitzer, R.L, & Williams, J.B.W. (2001).The PHQ-9 validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Long, H. A., French, D. P., & Brooks, J. M. (2020). Optimising the value of the critical appraisal skills programme (CASP) tool for quality appraisal in qualitative evidence synthesis. *Research Methods in Medicine & Health Sciences*, *1*(1), 31–42. https://doi.org/10.1177/2632084320947559

Mond, J., Hay, P.., Rodgers, B., Owen, C., & Beumont, P. J. (2004). Validity of the Eating Disorder Examination Questionnaire (EDE-Q) in screening for eating disorders in community samples. *Behaviour Research and Therapy*, *42*(5), 551–567. https://doi.org/10.1016/S0005-7967(03)00161-X

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine (1960)*, *166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092