**SUPPLEMENTARY MATERIAL**

**Supplementary Table 1: PRISMA checklist**

|  |  |  |  |
| --- | --- | --- | --- |
| **Section/topic** | **Item #** | **Checklist item** | **Reported on page #** |
| **TITLE** | | | |
| Title | 1 | Identify the report as a systematic review, meta-analysis, or both. | 1 |
| **ABSTRACT** | | | |
| Structured summary | 2 | Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number. | 3-4 |
| **INTRODUCTION** | | | |
| Rationale | 3 | Describe the rationale for the review in the context of what is already known. | 5 |
| Objectives | 4 | Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS). | 5-6 |
| **METHODS** | | | |
| Protocol and registration | 5 | Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number. | 6 |
| Eligibility criteria | 6 | Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale. | 6-7 |
| Information sources | 7 | Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched. | 6 |
| Search | 8 | Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated. | 6 |
| Study selection | 9 | State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis). | 6-7 |
| Data collection process | 10 | Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators. | 6-7 |
| Data items | 11 | List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made. | 6-9, S3 |
| Risk of bias in individual studies | 12 | Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis. | 7 |
| Summary measures | 13 | State the principal summary measures (e.g., risk ratio, difference in means). | 8-9 |
| Synthesis of results | 14 | Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I2) for each meta-analysis. | 8-9 |
| Risk of bias across studies | 15 | Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies). | 7 |
| Additional analyses | 16 | Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified. | 8-9 |
| **RESULTS** | | | |
| Study selection | 17 | Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram. | 10, Figure 1 |
| Study characteristics | 18 | For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations. | 10, Table 1 |
| Risk of bias within studies | 19 | Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12). | S10 |
| Results of individual studies | 20 | For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot. | Figures 2 & 3 |
| Synthesis of results | 21 | Present results of each meta-analysis done, including confidence intervals and measures of consistency. | 10-13 |
| Risk of bias across studies | 22 | Present results of any assessment of risk of bias across studies (see Item 15). | S12 |
| Additional analysis | 23 | Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]). | 10-13 |
| **DISCUSSION** | | | |
| Summary of evidence | 24 | Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers). | 13-18 |
| Limitations | 25 | Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias). | 17-18 |
| Conclusions | 26 | Provide a general interpretation of the results in the context of other evidence, and implications for future research. | 18 |
| **FUNDING** | | | |
| Funding | 27 | Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review. | 19 |

**Supplementary Table 2**: **Methods of Indexing Psychological Anorexia Nervosa Psychopathology**

Given the variable available methods for indexing psychological AN symptoms in the included studies, we devised a hierarchy whereby the most empirically supported method was preferred if available. This table illustrates this hierarchy, highlighting the order of preference, and the number of included studies utilizing each index of psychological AN symptomatology.

|  |  |  |
| --- | --- | --- |
| **Measure** | **Author** | **N** |
| Eating Disorder Examination | Cooper & Fairburn, 1987 | 10 |
| Eating Disorder Examination – Questionnaire | Fairburn & Beglin, 1994 | 2 |
| Eating Disorder Inventory | Garner et al., 1983 | 13 |
| Eating Attitudes Test | Garner et al., 1982 | 2 |
| YBOCS – ED | Sunday et al., 1995 | 1 |
| Short Evaluation of Eating Disorders – Anorexia Nervosa scale | Kordy, 2005 | 1 |
| EDE: Dietary Restraint scale | Cooper & Fairburn, 1987 | 3 |
| EDI-2: Drive for Thinness scale | Garner et al., 1983 | 3 |

**Supplementary Table 3: Random-effects model summaries**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| **Meta-analysis model** | **I2** | **τ2** | **Heterogeneity** | **Model results** | **Egger's test** |
| Weight at end of treatment | 17.6% | 0.01 | Q(34) = 38.6, *p* = 0.27 | *g* = 0.19, 95% CI (0.09, 0.29), *p* = 0.0002 | *p* = 0.93 |
| Cognition at end of treatment | 8.3% | 0.01 | Q(32) = 34.1, *p* = 0.37 | *g* = -0.02, 95% CI (-0.11, 0.08), *p* = 0.76 | *p* = 0.35 |
| Weight at follow up | 53.9% | 0.06 | Q(18) = 39, *p* = 0.003 | *g* = 0.14, 95% CI (-0.02, 0.3), *p* = 0.08 | *p* = 0.71 |
| Cognition at follow up | 30.7% | 0.02 | Q(18) = 30.4, *p* = 0.03 | *g* = -0.03, 95% CI (-0.15, 0.1), *p* = 0.69 | *p* = 0.12 |

Note: Summary effect sizes (and their confidence intervals) marginally differ from the more conservative cluster-robust estimates, which are corrected for

effect size dependencies.

**Supplementary Table 4: Moderator analyses**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Meta-analysis model** | **Year of publication** | **Illness duration** | **Age** | **Bias risk** | **Type of treatment** | **Follow-up time** | |
| Weight at EOT | *QM*(1) = 0.001, p = 0.98 | *QM*(1) = 0.28, p = 0.6 | *QM*(1) < 0.001, p = 0.99 | *QM*(1) = 0.51, p = 0.47 | *QM*(1) = 4.62, p = 0.03 | | — |
| Cognition at EOT | *QM*(1) 0.2, p = 0.65 | *QM*(1) = 1.92, p = 0.17 | *QM*(1) = 1.93, p = 0.16 | *QM*(1) = 0.41, p = 0.52 | *QM*(1) = 0.09, p = 0.76 | | — |
| Weight at follow-up | *QM*(1) = 0.36, p = 0.55 | *QM*(1) = 0.08, p = 0.78 | *QM*(1) < 0.001, p = 0.99 | *QM*(1) = 0.15, p = 0.7 | *QM*(1) = 6.5, p = 0.01 | | *QM*(1) = 0.02, p = 0.89 |
| Cognition at follow-up | *QM*(1) = 1.7, p = 0.19 | *QM*(1) = 3.19, p = 0.07 | *QM*(1) = 1.56, p = 0.21 | *QM*(1) = 1.06, p = 0.3 | *QM*(1) = 1.72, p = 0.19 | | *QM*(1) = 0.49, p = 0.48 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Meta-analysis model** | **Comparator category** | **Psychosocial category** | **Treatment platform** | **Psychosocial vs. pharmacological** | **Weight outcome** |
| Weight at EOT | *QM*(4) = 2.1, p = 0.72 | *QM*(2) = 2.49, p = 0.29 | *QM*(2) = 8.35, p = 0.02 | *QM*(1) = 0.43, p = 0.51 | *QM*(2) = 1.83, p = 0.4 |
| Cognition at EOT | *QM*(4) 6.76, p = 0.15 | *QM*(2) = 1.74, p = 0.42 | *QM*(2) = 1.09, p = 0.58 | *QM*(1) = 0.01, p = 0.94 | *QM*(2) = 0.27, p = 0.88 |
| Weight at follow-up | *QM*(3) = 2.75, p = 0.43 | *QM*(2) = 5.01, p = 0.08 | *QM*(2) = 6.34, p = 0.04 | *QM*(1) = 3.75, p = 0.053 | *QM*(2) = 3.99, p = 0.14 |
| Cognition at follow-up | *QM*(3) = 1.21, p = 0.75 | *QM*(2) = 0.16, p = 0.92 | *QM*(2) = 3.7, p = 0.16 | *QM*(1) = 0.03, p = 0.87 | *QM*(2) = 1.56, p = 0.46 |

**Supplementary Table 5: Cochrane risk of bias assessment for each eligible trial**

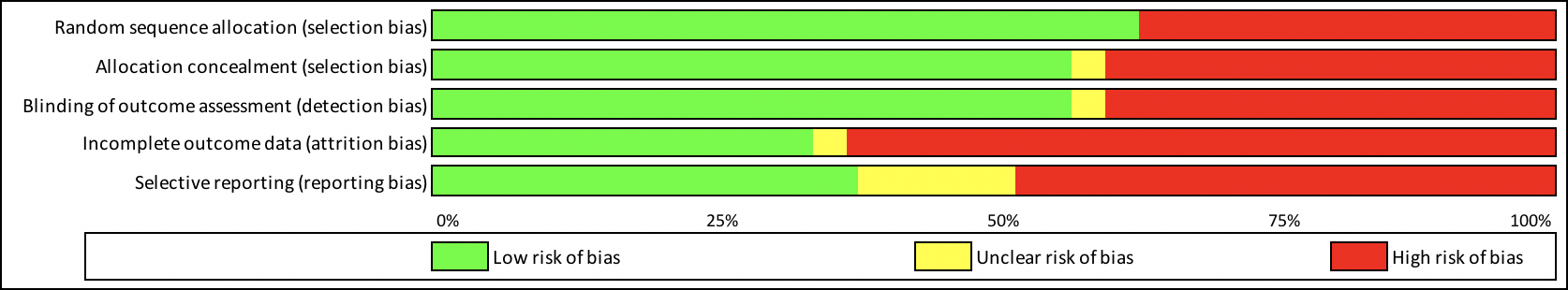
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Study** | **Random Sequence Allocation** | **Allocation Concealment** | **Blinding Outcome Assessment** | **Incomplete Assessment Data** | **Selective Reporting** |  |
| Attia et al (1998) |  |  |  |  |  |
| Robin et al (1999) |  |  |  |  |  |
| Eisler et al (2000) |  |  |  |  |  |
| Geist et al (2000 |  |  |  |  |  |
| Kaye et al (2001)20 |  |  |  |  |  |
| Fassino et al (2002)21 |  |  |  |  |  |
| Ball et al (2004)22 |  |  |  |  |  |
| McIntosh et al (2005)23 |  |  |  |  |  |
| Mondraty et al (2005)24 |  |  |  |  |  |
| Lock et al (2005)25 |  |  |  |  |  |
| Walsh et al (2006)26 |  |  |  |  |  |
| Brambilla et al (2007)27 |  |  |  |  |  |
| Gowers et al (2007)28 |  |  |  |  |  |
| Rigaud et al (2007)29 |  |  |  |  |  |
| Court et al (2010)30 |  |  |  |  |  |
| Lock et al (2010)31 |  |  |  |  |  |
| Attia et al (2011)32 |  |  |  |  |  |
| Hagman et al (2011)33 |  |  |  |  |  |
| Whitney et al (2011)34 |  |  |  |  |  |
| Schmidt et al (2012)35 |  |  |  |  |  |
| Godart et al (2012)36 |  |  |  |  |  |
| Powers et al (2012)37 |  |  |  |  |  |
| Touyz et al (2013)38 |  |  |  |  |  |
| Dalle Grave et al (2013)39 |  |  |  |  |  |
| Zipfel et al (2014)40 |  |  |  |  |  |
| Smith et al (2014)41 |  |  |  |  |  |
| Herpetz-Dahlmann et al (2014)42 |  |  |  |  | ? |
| Agras et al (2014)43 |  |  |  |  |  |
| Schmidt et al (2015)44 |  |  |  |  |  |
| Madden et al (2015)45 |  |  |  |  |  |
| Eisler et al (2016)46 |  |  |  |  |  |
| Le Grange et al (2016)47 | - | - | + | - | - |
| Parling et al (2016) |  |  |  |  |  |
| Herscovici et al (2017) |  |  |  |  |  |
| Russell et al (2018) |  |  |  |  |  |

****High risk of bias

Unclear risk of bias

Low risk of bias

**Supplementary Table 6: Cochrane risk of bias assessment summary for eligible trials**

****

**Supplementary Table 7**: **GRADE quality of evidence ratings for included studies, for both weight- and psychological AN symptoms.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Specialized treatments versus control treatments in the treatment of weight- and psychological symptoms of anorexia nervosa | | | | | | |
| **Patient population**: Adolescents and adults with anorexia nervosa  **Setting**: Inpatient, day program, and outpatient clinics  **Intervention**: Specialist psychosocial, medical, pharmacological and complementary/alternative treatments  **Comparison**: Treatment as usual and placebo groups | | | | | | |
| **Outcomes** | **No of participants at EOT (studies)** | **No of participants at follow-up (studies)** | **Relative effect at EOT (95% CI)** | **Relative effect at follow-up (95% CI)** | **Quality of evidence (GRADE)** | **Comments** |
| Weight-based AN symptom relief | 2,524 (35 RCTs) | 1,468 (19 RCTs) | *g* = 0·16 (0·05, 0·28) | *g* = 0·11 (-0·04, 0·27) | ⊕⊕ΟΟ  Low | Downgraded two levels due to (i) an elevated overall risk of bias, and (ii) some imprecision in overall findings stemming from several underpowered studies. Additional concerns relate to the fewer follow-up studies, and the unknown treatment status of almost all participants between EOT and follow-up. |
| Psychological AN symptom relief | 2,524 (35 RCTs) | 1,468 (19 RCTs) | *g =* -0·03 (-0·14, 0·08) | *g* = -0·001 (-0·11, 0·11) | ⊕⊕ΟΟ  Low | Downgraded two levels due to (i) an elevated overall risk of bias, and (ii) some imprecision in overall findings stemming from several underpowered studies. Additional concerns relate to the fewer follow-up studies, and the unknown treatment status of almost all participants between EOT and follow-up. Further concerns arise from the variety of indices used between studies to index psychological symptoms. |

**Supplementary Figure 1: The impact of participant age upon treatment outcomes**

Scatterplots demonstrating mean age plotted against effect sizes for weight (A) and psychological (B) outcomes at the end of treatment, and for weight (C) and psychological (D) outcomes at follow up. Larger points represent more precise effect sizes. The predicted effect size as a function of age with 95% confidence interval bounds is shown, which is derived from a mixed effects model. The respective moderator test result is also shown.

Figures/Figure%20age%20mod.pdf

**Supplementary Figure 2: The impact of publication year upon treatment outcomes**

Scatterplots demonstrating publication year plotted against effect sizes for weight (A) and psychological (B) outcomes at the end of treatment, and for weight (C) and psychological (D) outcomes at follow up. Larger points represent more precise effect sizes. The predicted effect size as a function of publication year with 95% confidence interval bounds is shown, which is derived from a mixed effects model. The respective moderator test result is also shown.

**Figures/Figure%20year%20mod.pdf**

**Supplementary Figure 3: The impact of treatment category upon treatment outcomes**

Uneven cell sizes across treatment category types in the present study precluded a meaningful comparison of psychosocial (*n* = 21), medical (*n* = 2), pharmacological (*n* =11) and complementary/alternative (*n* = 1) treatment interventions. As such, we compared psychosocial treatments against a conflated “All other treatments” category. This figure illustrates summary effect size estimates with 95% confidence intervals derived from random-effects meta-analysis estimates, for both weight- (A) and psychological (B) outcomes at EOT and follow-up.

Figures/treatment%20type.pdf

**Supplementary Figure 4: Contour-enhanced funnel plots**

Contour-enhanced funnel plots for weight (A) and psychological (B) outcomes at the end of treatment, and for weight (C) and psychological (D) outcomes at follow up. In these plots, effect sizes are plotted against standard errors to examine if statistically significant effects close to *p* = 0.05 are overrepresented in the included studies. Statistical significance can be calculated using a combination of effect size and standard error (assuming all studies used two-sided tests with a significance criterion of *p* = 0.05), and superimposed on the funnel plot. The light grey region captures *p* values between .1 and .05, and the dark grey region captures *p* values between .05 and .01. The white region corresponds to *p* values greater than .1. As there was no obvious overrepresentation of effect sizes lying within the light and dark grey regions, this is suggestive of a low likelihood of publication bias.

Figures/fig%20contour.pdf

**References**

Cooper Z, Fairburn CG. (1987). The Eating Disorder Examination: A semi-structured interview for the assessment of the specific psychopathology of eating disorders. International Journal of Eating Disorders, 6: 1-8.

Fairburn CG, Beglin SJ. (1994). Assessment of eating disorder psychopathology: Interview or self-report questionnaire? International Journal of Eating Disorders, 16; 363-370.

Garner Olmsted MP, Bohr Y, Garfinkel P. (1982). The Eating Attitudes Test: psychometric features and clinical correlates. Psychological Medicine, 12:871–878.

Garner DM, Olmsted M, Polivy J. (1983). Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia. International Journal of Eating Disorders, 2: 15-34.

Kordy H. (2005). Counting the COST: A European collaboration on the efficiency of psychotherapeutic treatment of patients with eating disorders. European Eating Disorders Review, 13, 153-158.

Sunday SR, Halmi KA, Einhorn A. (1995). The Yale-Brown-Cornell Eating Disorder Scale: A new scale to assess eating disorder symptomatology. International Journal of Eating Disorders, 18: 237-245.