Treatment outcomes for anorexia nervosa: a systematic review and meta-analysis of randomized controlled trials

Author for correspondence:
Stuart B. Murray, E-mail: stuart.murray@ucsf.edu, drstuartmurray@gmail.com

Stuart B. Murray[1], Daniel S. Quintana[2], Katharine L. Loeb[3], Scott Griffiths[4] and Daniel Le Grange[1,5]

[1]Department of Psychiatry, University of California, San Francisco, CA, USA; [2]NORMENT, KG Jebsen Centre for Psychosis Research, Division of Mental Health and Addiction, University of Oslo, and Oslo University Hospital, Oslo, Norway; [3]School of Psychology, Fairleigh Dickinson University, Teaneck, NJ, USA; [4]School of Psychology, University of Melbourne, Melbourne, VIC, Australia and [5]Department of Psychiatry and Behavioral Neuroscience, The University of Chicago, IL, USA

Abstract

Background. To determine the impact of specialized treatments, relative to comparator treatments, upon the weight and psychological symptoms of anorexia nervosa (AN) at end-of-treatment (EOT) and follow-up.

Methods. Randomized controlled trials (RCTs) between January 1980 and December 2017 that reported the effects of at least two treatments on AN were screened. Weight and psychological symptoms were analyzed separately for each study. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines were followed, and studies were assessed using the GRADE (Grading of Recommendations, Assessment, Development and Evaluation) criteria and Cochrane risk of bias tool.

Results. We identified 35 eligible RCTs, comprising data from 2524 patients. Meta-analyses revealed a significant treatment effect on weight outcomes at EOT [$g = 0.16$, 95% CI (0.05–0.28), $p = 0.006$], but not at follow-up [$g = 0.11$, 95% CI (-0.04 to 0.27), $p = 0.15$]. There was no significant treatment effect on psychological outcomes at either EOT [$g = -0.03$, 95% CI (-0.14 to 0.08), $p = 0.63$], or follow-up [$g = -0.001$, 95% CI (-0.11 to 0.11), $p = 0.98$]. There was no strong evidence of publication bias or significant moderator effects for illness duration, mean age, year of publication, comparator group category, or risk of bias (all p values > 0.05).

Conclusions. Current specialized treatments are more adept than comparator interventions at imparting change in weight-based AN symptoms at EOT, but not at follow-up. Specialized treatments confer no advantage over comparator interventions in terms of psychological symptoms. Future precision treatment efforts require a specific focus on the psychological symptoms of AN.

CAMBRIDGE
UNIVERSITY PRESS

Anorexia nervosa (AN) is a debilitating, often enduring, and potentially life-threatening psychiatric illness that is characterized by self-directed starvation, physical emaciation, an intense fear of weight gain, and a marked disturbance in how one's body shape and weight is experienced (American Psychiatric Association, 2013). AN is among the most lethal of all psychiatric presentations, demonstrating a crude mortality rate of 5.6% per decade (Arcelus *et al.*, 2011), and a 57-fold increased risk for suicidality (Keel *et al.*, 2003). Even in non-lethal presentations, AN ranks as the third leading cause of chronic illness in adolescents (Matthews *et al.*, 2011), and imparts an array of multi-systemic organ damage, including cardiac abnormalities, structural and functional brain impairment, and early-onset bone disease (Mitchell and Crow, 2006). Importantly, illness duration may extend to over 20 years for more than half of those afflicted (Fichter *et al.*, 2017), highlighting marked chronicity. Efficacious treatment for AN is therefore of high importance.

However, even the most promising interventions leave more than half of those treated unremitted (Watson and Bulik, 2013). A comprehensive review at the turn of the twenty-first century concluded that treatment outcomes for AN had not improved over the preceding 50 years (Steinhaussen, 2002), and a call has been put forth to temporarily suspend large-scale randomized controlled trials (RCTs) until novel interventions are developed (Fairburn, 2005). A clear explication of core illness mechanisms, and their response to treatment, is a fundamental prerequisite for the development of precision treatments for AN. Elucidating the disconnect between weight and psychological outcomes in the treatment of AN speaks directly to this mission. However, no meta-analysis has been conducted to delineate between these potentially discrepant symptom dimensions in the treatment of AN. Our objective here was to conduct a systematic review and meta-analysis of recent RCTs that assessed the efficacy of treatments for

AN, when indexed across both weight and psychological symptoms, at post-intervention and, where available, longer term follow-up.

Methods

Selection procedures

A systematic review was performed according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Moher *et al.*, 2009) (Supplementary Table S1), and the protocol for this review was prospectively registered with PROSPERO (CRD42016049831). Two authors (SBM and SG) independently searched MEDLINE/PubMed, PsychINFO, ScienceDirect, and EMBASE databases from January 1980 through to December 2017. Key search terms included relevant combinations of *eating disorders*, *anorexia nervosa*, *treatment*, *treatment outcome*, *outcome*, and *trial*, with Boolean operators. Studies in any language were considered, although all included RCTs were published in English. The search was completed with an additional screening of reference lists of eligible trials, screening existing systematic and narrative reviews for AN, and a manual journal search. Abstracts were screened and the full texts of relevant studies were retrieved. If full texts were not available, corresponding authors were contacted and full texts requested. Two authors (SBM and SG) selected the final studies for inclusion.

Selection criteria

Studies were included if they (i) were RCTs published between January 1980 and December 2017, (ii) reported the effect of at least one specialized treatment and one comparator or control treatment on patients with a diagnosis of AN, across at least two time points (i.e. pre- and post-treatment), and (iii) reported indices of both weight and psychological symptoms. For all studies, the designation of specialized *v.* comparator status was directly informed by the original studies. We did not assess data relating to mortality, since this was recently explicated (Arcelus *et al.*, 2011). Studies involving participants without core AN-like psychological criteria (e.g. 'non-fat-phobic' AN) were excluded, as these clinical profiles would confound outcomes related to the core diagnostic criteria. We reviewed all RCTs reflecting these criteria and assessed the methodological quality of studies according to GRADE (Grading of Recommendations, Assessment, Development and Evaluation) guidelines (Guyatt *et al.*, 2011). Study participants of any age were eligible, although given the widely reported differential prognoses (Steinhaussen, 2002; Watson and Bulik, 2013) for adolescent *v.* adult AN, respectively, we treated mean age as a potential moderator variable. Treatment types included in the analyses were psychosocial, pharmacological, medical, and complementary/alternative interventions.

Risk of bias and data extraction

The Cochrane risk of bias tool (Higgins *et al.*, 2011) was used to evaluate risk of bias. Performance bias was not assessed due to non-feasibility of blinding therapists and patients in RCT designs for psychosocial treatments. Studies without a clearly described method of blinding outcome assessors were rated as high risk of detection bias. Studies not describing an established method of participant randomization, or the concealment of this random allocation, were rated as high risk of random sequence generation bias, and selection bias, respectively. Likewise, studies not describing a protocol for intent-to-treat analyses were rated as high risk of attrition bias. Studies reporting analyses that diverged from prospective trial protocols, or where important data were not reported, were rated as high risk of reporting bias. For use in meta-regression analyses, we computed an overall risk of bias score for each study by awarding one point for each potential source of bias rated as low risk.

Statistical analysis

To delineate treatment outcomes, we indexed the findings of each study according to both weight and psychological symptom categories. In light of the inter-trial variability in the measures used for indexing psychological AN symptoms, a hierarchy was developed whereby the most empirically supported method in each study was preferred (Supplementary Table S2). All measures indexed core aspects of cognitive AN psychopathology, including global eating disorder symptom severity, dietary restraint, the drive for thinness, and food-related obsessionality. Statistical analysis was performed with R statistical software 3.3.2, using the *metafor* package. The dataset and script to perform the analyses are available at https://osf.io/q7v2d/?view_only=c3cdaf346298411eab9ed15e863c9f21. The primary outcomes of interest were the effect of treatment on both weight status and psychological AN symptomatology at the end-of-treatment (EOT) and at follow-up. Multiple effect sizes derived from the same study (e.g. weight and psychological outcomes) are statistically dependent, forming clusters of internally correlated effect size estimates. If effect size covariance is known, this can be included in models to adjust for these statistically dependent clusters. However, without access to original datasets, the covariance between effect size estimates of included studies is rarely available, as covariances are seldom reported. One approach is to assume a fixed covariance value; however, inaccuracy can lead to errors in effect size estimation (Cuijpers *et al.*, 2017). Cluster-robust meta-analyses can account for statistically dependent clusters without assuming a fixed covariance value when covariances are not reported (Hedges *et al.*, 2010). As such, cluster-robust meta-analysis was applied to assess the primary outcomes. A random mixed-effects meta-analysis assuming a diagonal *v.* matrix was used to construct cluster-robust models for each outcome. Contrasts were then performed to assess for differences in effect sizes between weight and psychological outcomes at each point, and any changes in effect sizes over time from EOT to follow-up. Cluster-robust inferences were also used to assess differences in dropout frequencies between treatments, and whether this effect changed between EOT and follow-up. For ease of interpretation, weight outcomes (whereby a mean *increase* is a positive outcome) and psychological symptom outcomes (whereby a mean *decrease* is a positive outcome) were adjusted, such that positive values in both domains represent symptom improvement.

Secondary analysis was performed on weight and psychological outcomes at each time point separately to estimate the impact of the potential moderators: illness duration, treatment platform (inpatient, outpatient, mixed level of care), year of publication, age, risk of bias, type of weight outcome measure, and follow-up length. Owing to the greater number of psychosocial treatment trials ($n = 21$) *v.* medication ($n = 11$), medical ($n = 2$), and complementary/alternative ($n = 1$) treatment trials, the moderating effect of treatment type was conceptualized as a function

of psychosocial treatments *v.* other treatments. As a supplementary analysis, we also conceptualized treatment type as a function of psychosocial treatments *v.* medication treatments. Relatedly, in delineating the potentially moderating effect of the various psychosocial interventions included, psychosocial treatment type (family therapy, individual therapy, group treatment) was also assessed as a potential moderator for psychosocial treatment trials. Lastly, due to variability in comparator interventions employed in trials (placebo, active treatment, non-specific treatment, treatment-as-usual), the moderating effect of comparator intervention type was also assessed. Small study bias, which includes both publication bias and study quality bias (Egger *et al.*, 1997), was assessed by performing Egger's regression test (Egger *et al.*, 1997). Contour-enhanced funnel plots, which superimpose key areas of statistical significance ($p = 0.1$, $p = 0.05$, $p = 0.01$), were constructed to specifically assess for the risk of publication bias (Peters *et al.*, 2008). Effect sizes from multi-arm trials ($n = 4$) were merged for these secondary analyses, as recommended for random-effects models (Rücker *et al.*, 2017).

Results

We identified 35 eligible RCTs, which included a total of 2524 participants (Fig. 1). Eligible RCTs included studies with adolescent populations ($n = 9$), in which the mean age was 15.01 years (S.D. = 0.39); adult populations ($n = 8$), in which the mean age was 27.1 years (S.D. = 4.36); and mixed populations ($n = 18$), in which the mean age was 21.86 years (S.D. = 4.56) (Table 1). Four RCTs yielded two effect sizes for each outcome, and one study featured a previously tested experimental treatment as a comparator treatment. Individual study effects at EOT are presented in Fig. 2, and at follow-up in Fig. 3.

For overall treatment effects, a summary of study effect sizes and variances is presented in Fig. 4. Cluster-robust models revealed a significant treatment effect on weight outcomes at the EOT [$g = 0.16$, 95% CI (0.05–0.28), $p = 0.006$]. There was no significant treatment effect on psychological outcomes at EOT [$g = -0.03$, 95% CI (-0.14 to 0.08), $p = 0.63$]. At follow-up, there was no significant treatment effect upon weight outcomes [$g = 0.11$, 95% CI (-0.04 to 0.27), $p = 0.15$] or on psychological outcomes [$g = -0.001$, 95% CI (-0.11 to 0.11), $p = 0.98$]. Heterogeneity between the four models was on the border of statistical significance [$F_{(4,31)} = 2.5$, $p = 0.06$]. Contrasts showed that the effects of treatment on weight outcomes at EOT were larger than the psychological outcomes [$g = 0.19$, S.E. = 0.08; $F_{(1,31)} = 5.77$, $p = 0.02$]. There was no significant difference between weight and psychological outcomes at follow-up [$F_{(1,31)} = 2.1$, $p = 0.16$]. There was no significant difference in the effect of treatment between EOT and follow-up on psychological outcomes [$F_{(1,31)} = 0.18$, $p = 0.68$] or weight outcomes [$F_{(1,31)} = 0.92$, $p = 0.35$]. Cluster-robust models indicated no significant treatment effect on drop-out frequency between treatment and comparator groups at EOT [log odds ratio estimate = -0.01, 95% CI (-0.25 to 0.24), $p = 0.97$] or at follow-up [log odds ratio estimate = 0.04, 95% CI (-0.32 to 0.44), $p = 0.81$]. There also was no significant difference in drop-out frequency effects between EOT and follow-up [$F_{(1,33)} = 0.07$, $p = 0.79$].

For secondary analysis, separate random-effects models of weight and psychological outcomes at EOT and follow-up were performed to assess the effect of potential moderators (Supplementary Table S3). Weight outcomes [$Q(18) = 39$, $p = 0.003$] and psychological [$Q(18) = 30.4$, $p = 0.03$] outcomes
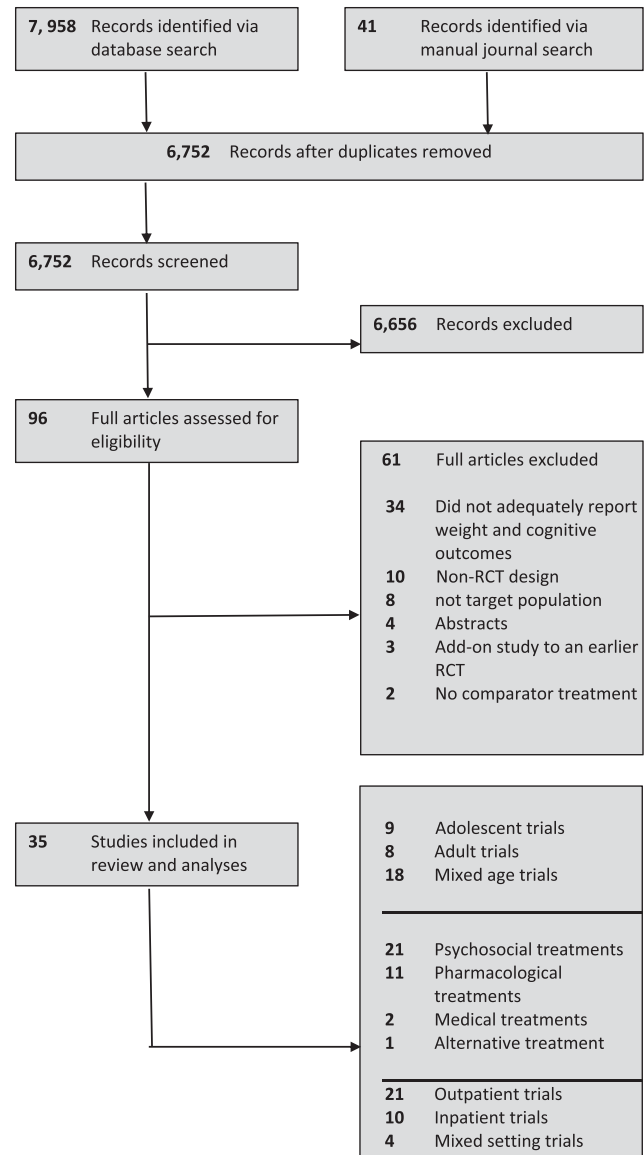

Fig. 1. PRISMA flowchart of study selection.

at follow-up displayed significant heterogeneity. There were no significant moderating effects of illness duration, mean age, year of publication, risk of bias, type of weight outcome, or follow-up time for any of the models (Supplementary Table S4; Supplementary Figs S1 and S2). The moderating effect of treatment type on weight outcomes at EOT was statistically significant [$Q_M(1) = 4.62$, $p = 0.03$], with psychosocial treatments, relative to their comparator intervention groups [$g = 0.13$, S.E. = 0.05, $p = 0.01$], yielding a significantly smaller summary effect size ($z = -2.15$, $p = 0.03$) than the other specialized treatment comparisons combined ($g = 0.36$, S.E. = 0.1, $p = 0.0001$) (Supplementary Fig. S3). When only comparing psychosocial and pharmacological treatments, there was no statistically significant moderating effect of treatment on weight or psychological outcomes at EOT or follow-up (Supplementary Table S4). There was a statistically significant effect of treatment type on weight outcomes at follow-up [$Q_M(1) = 6.5$, $p = 0.01$], with psychosocial treatment comparisons ($g = 0.07$, S.E. = 0.08, $p = 0.35$) yielding a significantly smaller

Table 1. A summary of eligible studies

	Population	Treatment	Comparator treatment	N treated in specialty treatment	N treated in comparator treatment
Attia et al. (1998)	Adolescents and adults aged 16–45	Fluoxetine	Placebo	15	16
Robin et al. (1999)	Adolescents and young adults aged 11–20 years	Behavioral family systems therapy	Ego-oriented individual therapy	19	18
Eisler et al. (2000)	Adolescents aged 12–18 years	Separated family therapy	Conjoint family therapy	21	19
Geist et al. (2000)	Adolescents aged 12–17 years	Family therapy	Family psychoeducation	12	13
Kaye et al. (2001) a	Adults aged 19+ years	Fluoxetine completers	Placebo	10	3
Kaye et al. (2001) b		Fluoxetine non-completers	Placebo	6	16
Fassino et al. (2002)	Adolescents and adults aged 16–35 years	Citalopram	Wait list control	26	26
Ball et al. (2004)	Adolescents and young adults aged 13–23 years	Cognitive behavioral therapy	Behavioral family therapy	13	1
McIntosh et al. (2005) a	Adolescents and adults aged 17–40 years	Cognitive behavioral therapy	Non-specific supportive clinical management	19	16
McIntosh et al. (2005) b		Interpersonal psychotherapy	Non-specific supportive clinical management	21	16
Mondraty et al. (2005)	Adults	Olanzapine	Chlorpromazine	8	7
Lock et al. (2005)	Adolescents aged 12–18 years	6-month family-based treatment	12-month family-based treatment	44	42
Walsh et al. (2006)	Adolescents and adults aged 16–45 years	Fluoxetine	Placebo	49	44
Brambilla et al. (2007)	Adult aged 19+	Olanzapine	Placebo	15	15
Gowers et al. (2007) a	Adolescents aged 12–18 years	Inpatient care	General outpatient care	57	55
Gowers et al. (2007) b		Specialist outpatient care	General outpatient care	55	55
Rigaud et al. (2007)	Adults	Cyclic enteral nutrition	Control re-feeding	41	40
Court et al. (2010)	Adolescents and adults	Quetiapine	Treatment as usual	15	18
Lock et al. (2010)	Adolescents aged 12–18 years	Family-based treatment	Adolescent-focused therapy	61	60
Attia et al. (2011)	Adolescents and adults aged 16+ years	Olanzapine	Placebo	11	12
Hagman et al. (2011)	Adolescents and young adults aged 12–21	Risperidone	Placebo	18	22
Whitney et al. (2011)	Participant age not reported	Family day workshops	Family therapy	25	23
Schmidt et al. (2012)	Adolescents and adults aged 18+ years	Maudsley model of anorexia nervosa treatment for adults	Specialist supportive clinical management	34	37
Godart et al. (2012)	Adolescents and young adults aged 13–19	Family therapy	Treatment as usual	30	30
Powers et al. (2012)	Adults aged 19–65 years	Quetiapine	Placebo	6	9
Touyz et al. (2013)	Adults aged 19+ years	Cognitive behavioral therapy	Specialist supportive clinical management	31	32
Dalle-Grave et al. (2013)	Adolescents and adults aged 14–65 years	Cognitive behavioral therapy – broad	Cognitive behavioral therapy – focused	38	42
Zipfel et al. (2014) a	Adults aged 19+ years	Focal dynamic psychotherapy	Optimized treatment as usual	80	82
Zipfel et al. (2014) b		Cognitive behavioral therapy	Optimized treatment as usual	80	82

(Continued)

Table 1. (*Continued.*)

|	Population	Treatment	Comparator treatment	N treated in specialty treatment	N treated in comparator treatment
Smith et al. (2014)	Adolescents and adults aged 16+ years	Acupuncture	Acupressure	13	13
Herpetz-Dahlmann et al. (2014)	Adolescents aged 11–18 years	Day-patient treatment	Inpatient treatment	87	85
Agras et al. (2014)	Adolescents aged 12–18 years	Family-based treatment	Systemic family therapy	78	80
Schmidt et al. (2015)	Adolescents and adults aged 18+	Maudsley model of anorexia nervosa treatment for adults	Specialist supportive clinical management	72	70
Madden et al. (2015)	Adolescents aged 12–18 years	Inpatient medical stabilization	Inpatient weight restoration	41	41
Eisler et al. (2016)	Adolescents and young adults aged 13–20 years	Multi-family therapy	Family-based treatment	86	83
Le Grange et al. (2016)	Adolescents aged 12–18 years	Parent-focused therapy	Family-based treatment	52	55
Parling et al. (2016)	Adults aged 18–51	Acceptance and commitment therapy	Treatment as usual	24	19
Herscovici et al. (2017)	Adolescents and young adults aged 12–20 years	Family therapy with meal session	Family therapy with no meal session	11	12
Russell et al. (2018)	Adolescents and adults aged 16–60 years	Intranasal oxytocin	Placebo	20	21
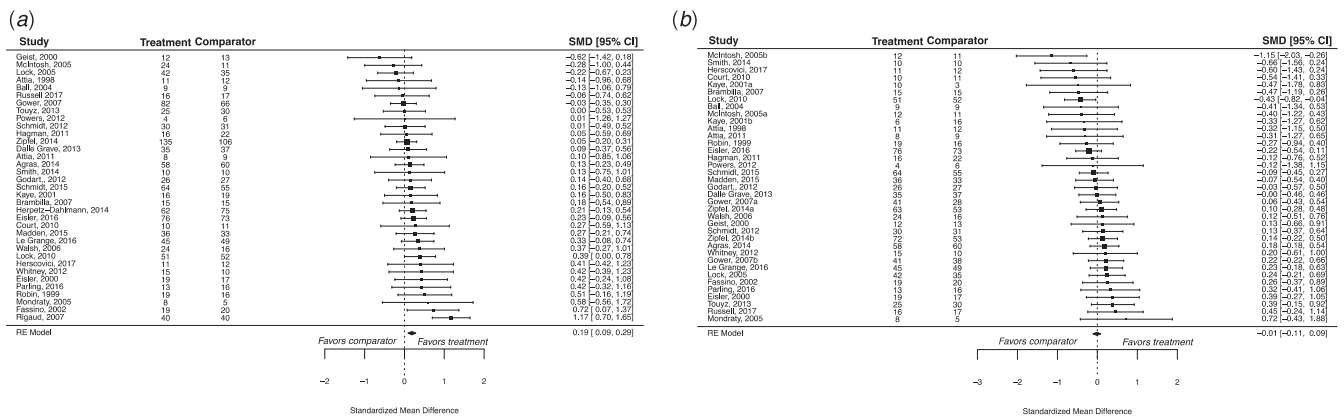
Fig. 2. A summary of individual study effects at EOT. The comparison between specialized and comparator interventions at EOT for weight (*a*) and psychological (*b*) outcomes, derived from separate random-effects models. Hedges' *g* point estimates are depicted by filled squares and the filled diamonds reflect the estimated summary effect sizes. Error bars and diamond widths represent 95% confidence intervals. Positive values represent effects favoring the specialized treatment, whereas negative values represent effects favoring the comparator treatment. Participant numbers refer to those completing treatment. Note that summary effect sizes (and confidence intervals) marginally differ from the more conservative cluster-robust estimates, which corrected for effect size dependencies.

relative summary effect size ($z = -2.55$, $p = 0.01$) than the other specialized treatment comparisons combined ($g = 0.6$, S.E. = 0.19, $p = 0.002$). The moderating effect of psychosocial treatment category (family therapy treatment $N = 10$, individual treatment $N = 9$, group treatment = 2) on weight outcomes at follow-up was also on the border of statistical significance [$Q_M(2) = 5.01$, $p = 0.08$]. Only the summary effect size for family therapy treatments ($k = 10$) was statistically significant ($g = 0.25$, S.E. = 0.09, $p = 0.007$). There was no statistically significant moderator effect of comparator category at EOT for weight [$Q_M(4) = 2.1$, $p = 0.72$] or psychological [$Q_M(4) = 6.76$, $p = 0.15$] outcomes, or at

follow-up for weight [$Q_M(3) = 2.75$, $p = 0.43$] or psychological [$Q_M(3) = 1.21$, $p = 0.75$] outcomes (Fig. 5).

There was a statistically significant moderating effect of treatment platform on weight outcomes at EOT [$Q_M(2) = 8.35$, $p = 0.02$]. The summary effect sizes for inpatient ($g = 0.6$, $p < 0.001$) and outpatient ($g = 0.16$, $p = 0.004$) platforms for weight at EOT were statistically significant. The summary weight outcome for inpatients was significantly larger than the mixed ($z = 2.83$, $p = 0.005$) and outpatient ($z = 2.68$, $p = 0.007$) outcomes. There was also a statistically significant moderating effect of treatment platform on weight outcomes at follow-up [$Q_M(2) = 6.34$, $p = 0.04$].
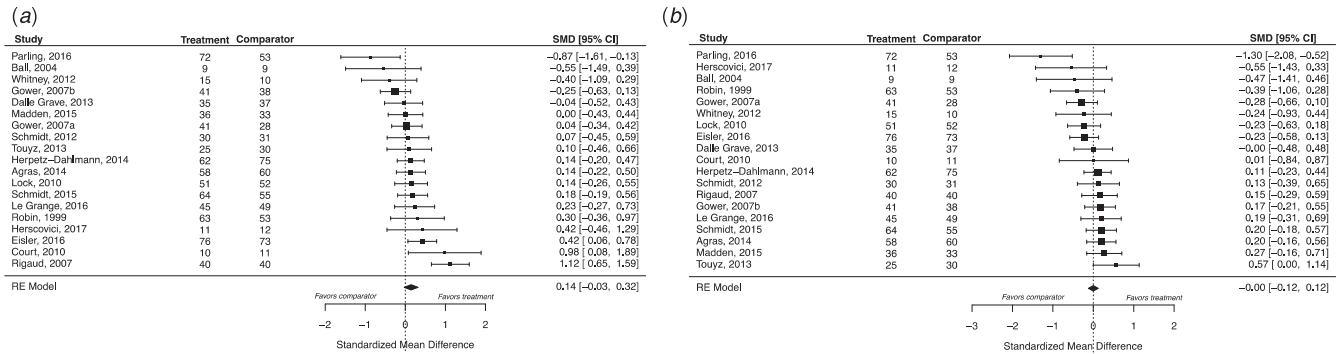
(a)

(b)

Fig. 3. A summary of individual study effects at follow-up. The comparison between specialized and comparator interventions at follow-up for weight (*a*) and psychological (*b*) outcomes, derived from separate random-effects models. Hedges' *g* point estimates are depicted by filled squares and the filled diamonds reflect the estimated summary effect sizes. Error bars and diamond widths represent 95% confidence intervals. Positive values represent effects favoring the specialized treatment, whereas negative values represent effects favoring the comparator treatment. Participant numbers refer to those completing treatment. Note that summary effect sizes (and confidence intervals) marginally differ from the more conservative cluster-robust estimates, which corrected for effect size dependencies.
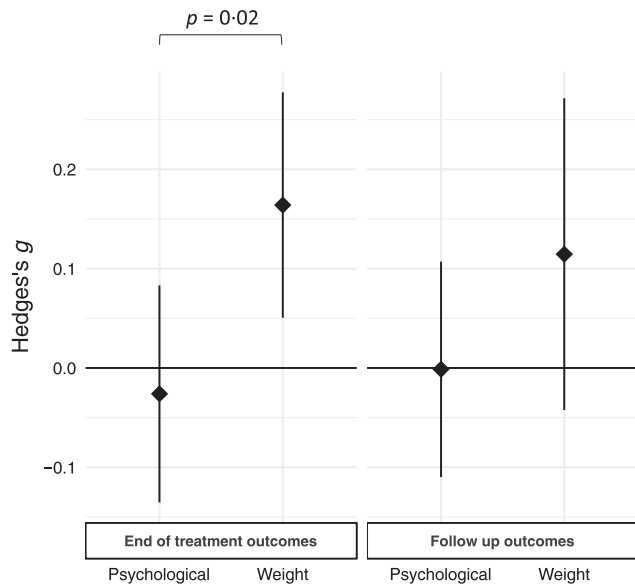

Fig. 4. Summary effects of treatments on weight and psychological symptoms at EOT and follow-up. Summary effect size estimates are illustrated with 95% confidence intervals derived from cluster-robust meta-analysis, for the effects of treatment upon weight and psychological symptoms at EOT and follow-up, respectively. Positive values represent effects favoring the active treatment, whereas negative values represent effects favoring the comparator treatment. The effect of treatment on weight symptoms at EOT was significantly larger than the effect on psychological symptoms.

For weight at follow-up, the summary effect size for inpatient studies was statistically significant ($g = 0.53$, $p = 0.01$) and the summary effect size for outpatient studies was on the border of statistical significance ($g = 0.17$, $p = 0.06$). The summary weight outcome for inpatients was significantly larger than mixed treatment platform outcomes ($z = 2.47$, $p = 0.01$). There was no evidence of small study bias for the weight and psychological outcome EOT and follow-up models (Egger's regression test *p* values > 0.05; Supplementary Table S3). Inspection of contour-enhanced funnel plots revealed no over-representation of effect sizes in the significance contours (Supplementary Fig. S4).

Re-analyzing the primary meta-analysis after removing studies comparing differential doses of the same treatment or nuanced

forms of the same treatment (Lock *et al.*, 2005; Dalle Grave *et al.*, 2013; Herscovici *et al.*, 2017) revealed similar results. Cluster-robust models revealed a significant treatment effect on weight outcomes at the EOT [$g = 0.18$, 95% CI (0.06–0.3), $p = 0.006$]. There was no significant treatment effect on psychological outcomes at EOT [$g = -0.01$, 95% CI (-0.13 to 0.11), $p = 0.86$]. At follow-up, there was no significant treatment effect upon weight outcomes [$g = 0.12$, 95% CI (-0.05 to 0.3), $p = 0.15$] or on psychological outcomes [$g = 0.01$, 95% CI (-0.11 to 0.12), $p = 0.91$]. Heterogeneity between the four models was on the border of statistical significance [$F_{(4,28)} = 2.6$, $p = 0.06$]. Contrasts showed that the effects of treatment on weight outcomes at EOT were larger than the psychological outcomes [$g = 0.19$, S.E. = 0.09; $F_{(1,28)} = 4.79$, $p = 0.04$]. There was no significant difference between weight and psychological outcomes at follow-up [$F_{(1,28)} = 1.89$, $p = 0.18$]. There was no significant difference in the effect of treatment between EOT and follow-up on psychological outcomes [$F_{(1,28)} = 0.91$, $p = 0.35$] or weight outcomes [$F_{(1,28)} = 0.07$, $p = 0.79$].

Cochrane risk of bias assessment found that 37% of the included studies had a high risk of random sequence generation bias, 40% had a high risk of allocation concealment bias, 40% had a high risk of detection bias, 63% had a high risk of attrition bias, and 48% had a high risk of reporting bias (Supplementary Tables S5 and S6). Overall, the methodological quality of the included studies was deemed to be low (Supplementary Table S7).

Discussion

This systematic review and meta-analysis represents a comprehensive synthesis of empirical data relating to the treatment of AN, and is the first to grade the methodological quality and risk of bias across RCTs for AN. Importantly, we delineated weight and psychological symptoms, and integrated data relating to psychosocial, pharmacological, medical, and complementary/alternative treatments for AN, all from RCTs, and included all patient age groups. Most trials meeting inclusion criteria were from the year 2000 onwards, when reporting psychological outcomes alongside weight outcomes became more commonplace. Overall, we found that specialized treatments conferred a significant treatment effect over and above comparator treatments in terms of
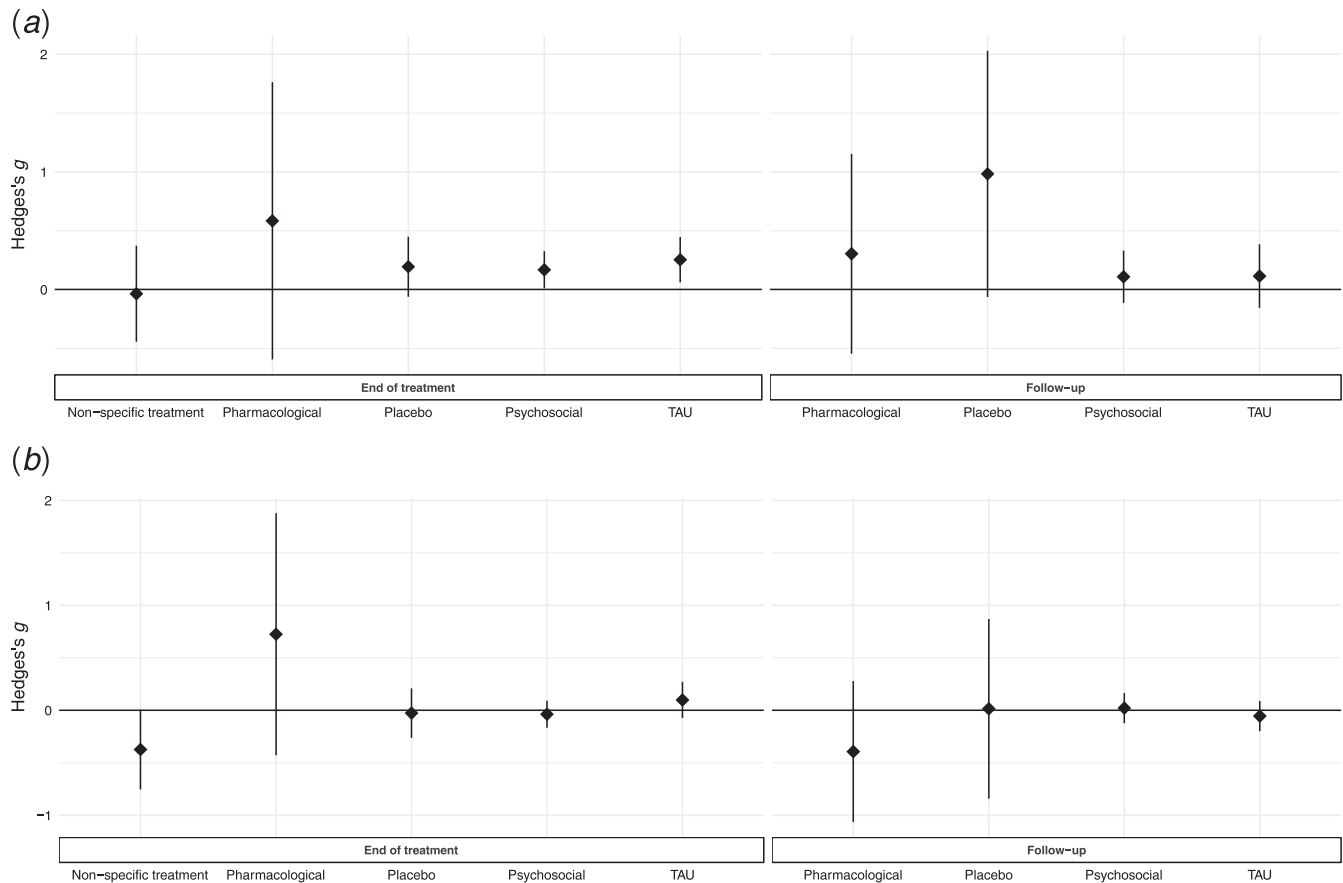
Fig. 5. Summary effects of comparator group categories on weight and psychological symptoms at EOT and follow-up. Summary effect size estimates with 95% confidence intervals derived from random-effects meta-analysis estimates for compactor group categories on weight (*a*) and psychological (*b*) symptoms at EOT and follow-up. Positive values represent effects favoring the active treatment, whereas negative values represent effects favoring the comparator treatment. TAU = treatment as usual.

weight-based symptom improvement at EOT, but not at follow-up. In terms of psychological symptom change, specialized treatments did not confer any advantage over comparator treatments at EOT or follow-up. These data indicate a distinct and discrepant trajectory of weight *v.* psychological AN symptoms throughout treatment, and highlight that current specialized treatments are more adept at altering the course of weight-based AN symptomatology, as opposed to psychological AN symptomatology.

The relatively greater impact of specialized treatments over comparator interventions upon weight-based outcomes at EOT is an important finding when considering the likelihood of starvation-related death in AN, as well as the striking cost of inpatient weight restoration in AN, which in the USA may total up to $321 300 per patient (Guarda *et al.*, 2017). As such, this incremental benefit of specialized interventions in weight-based symptoms confers both clinical and economic advantages. However, this change in weight-based AN symptomatology throughout specialized treatments did not parallel a corresponding shift in psychological AN symptoms; no beneficial effects of these treatments were found in psychological AN symptomatology at EOT, over and above what was demonstrated in comparator treatments. Importantly, psychological symptoms are thought to be central maintaining mechanisms in AN psychopathology (Fairburn *et al.*, 1999; DuBois *et al.*, 2017), and this cluster of cognitions (e.g. overvaluation of shape and weight, fear of weight gain) is hypothesized to drive the behavioral features of AN (e.g. dietary restriction), which in turn drive weight loss.

Moreover, with psychological symptoms being a key precursor to treatment dropout (Woodside *et al.*, 2004) and relapse in AN (Keel *et al.*, 2005), persistent psychological symptomatology, even in the context of improvements in weight-based symptoms, represents a detriment to long-term prognosis.

Relatedly, findings at follow-up suggest no sustained augmentative effect of specialized treatments in terms of weight-based symptom shift, relative to comparator treatments. It is unclear whether this represents a weight loss in specialized treatments between EOT and follow-up, or a greater improvement in weight in comparator treatments between EOT and follow-up, or both, and this important question requires further study. Similarly, no augmentative effect of specialized treatments upon the psychological symptoms of AN was noted at follow-up, relative to comparator treatments. These findings appear to be robust, and extend across all types of control interventions used in the comparisons. Cumulatively, these data suggest that current specialized treatments for AN, while more adept at improving patient weight in the shorter term, are not more effective than comparator treatments in producing sustained benefits in weight status, or in curtailing the psychological symptoms of AN at EOT or follow-up. An important caveat relating to follow-up findings, however, rises from the ambiguity of treatment status beyond EOT in most RCTs. Of all included studies, only about half provided follow-up data, and among those, only one study reported information relating to ongoing treatment engagement across groups beyond EOT. An ethical obligation facing researchers upon the

completion of RCTs, in the event of ongoing clinically significant symptoms, relates to the provision of, or referral to, continued care. However, it is unclear in the included RCTs whether those participants requiring ongoing treatment beyond EOT actually engaged in this, and if so, whether the type of follow-up care was similar or discrepant to the treatment received during the trial. The scarcity of these data compromises the interpretation of results beyond EOT, and potentially minimizes differences over the long term. Notwithstanding, these findings raise important questions as to the utility of current specialized treatments in their capacity to allay the psychological symptoms of AN in the short or longer term.

In terms of treatment setting, findings suggest that specialty treatments in mixed treatment settings (i.e. partial hospital programs) do not reliably yield differentiated weight-based symptom outcomes from their comparator treatments, as opposed to specialty treatments in inpatient and outpatient settings, at both EOT and follow-up. This is consistent with the literature suggesting a less well-developed theoretical base and clinical guidelines for treatment in such settings, and the noted difficulty in isolating the active components of treatments over such a protracted duration of treatment exposure (i.e. up to 10 h per day) (Freidman *et al.*, 2016). With respect to treatment category, findings suggest that the effects of psychosocial interventions were in fact *smaller* than treatment effects of other interventions in imparting weight-based symptom change, both at EOT and follow-up. Moreover, treatment category did not moderate the non-significant treatment effects on psychological symptoms at either EOT or follow-up. Among psychosocial treatments, a *marginally* greater treatment effect on weight outcomes at EOT was demonstrated for family-based therapies, which is consistent with the largely weight-driven focus in these treatments, although this effect did not extend to psychological outcomes. Cumulatively, these results imply that overall, interventions that address purported psychological mechanisms of AN in the most face-valid manner are failing to demonstrate treatment effects over and above comparator treatments. This underscores the need to explicate illness mechanisms and target them in novel, precision interventions.

We found no moderating effect of publication year upon both weight and psychological outcomes, suggesting that specialized treatments are not incrementally improving outcomes, relative to comparator treatments, over time. This is consistent with the platform put forth over a decade ago, that a plateau in treatment efficacy for AN has long been reached (Steinhaussen, 2002; Fairburn, 2005). These findings underscore the pressing need to elucidate AN mechanisms and develop corresponding precision treatments that may advance clinical outcomes. Additionally, the finding that age, as well as its correlate, illness duration, did not moderate treatment outcome is noteworthy. Current theorizing postulates that prognosis is more favorable for younger AN presentations, and that treatments targeted at adolescents are more efficacious than those targeted at adults (Watson and Bulik, 2013). However, our findings suggest that adolescents engaging in specialist treatments are not more likely to differentially outperform their counterparts engaging in comparator treatments, relative to adults. Thus, adolescents' better prognosis is not due to specific specialist treatment effects *per se*, since the difference between specialist and comparator groups is comparable in both adolescents and adults, and therefore more likely reflects a pervasive prognostic favorability of this age group.

An important challenge in the treatment of AN relates to the high rates of treatment dropout, which typically ranges from 20%

to 46% (DeJong *et al.*, 2012). The present study suggests that specialized treatments do not appear to confer any advantages over comparator treatments in treatment acceptability, as expressed by rate of attrition. As such, improving treatment tolerability and patient retention remains an important goal in future treatment development efforts.

Limitations to this meta-analysis include the low methodological quality of included studies, and the elevated risk of bias. However, risk of bias did not emerge as a moderator of treatment effects, and there was no evidence of publication bias. The greatest risk of bias stemmed from inadequately accounting for patient attrition during analyses, and selective reporting, suggesting a greater need for transparency in treatment trials. Moreover, the challenges of conducting RCTs in the context of AN are well known, and inherently extend to conducting meta-analyses in this domain. For instance, the medical complexities which arise rapidly in the context of AN necessitate ethical obligations mandating against waitlist-controlled treatment trials or the withholding of active treatment, and as such, psychosocial treatments have necessarily been compared with a heterogeneous array of active treatments during controlled trials. Heterogeneity in existing treatment approaches compounds this challenge, and while we statistically controlled for the effect of variability in both comparator and specialty treatment type in moderator analyses, this heterogeneity is noteworthy. Lastly, it should be noted that these findings reflect outcomes from RCTs in which both weight and psychological outcomes were reported, and may not reflect findings from studies which did not meet inclusion criteria. These factors ought to be carefully examined in interpreting results in treatment trials in these populations.

Conclusions

Cumulatively, our results point toward discrepant symptom pathways for weight *v.* psychological symptom shift throughout treatment for AN, which should be independently indexed in treatment trials. Moreover, the absence of a moderating effect of age and illness duration upon these discrepant pathways points toward a robustness of these pathways, irrespective of such patient characteristics. Specialized psychosocial treatments appear less able to alter the course of weight-based symptoms in AN, relative to other specialized treatments, although all specialized treatments do not appear to additively alter the course of psychological symptomatology above what is demonstrated in treatment as usual, control, or placebo groups. More broadly, these results question the utility of focusing on weight-based symptom remission as the terminal goal of treatment in AN. Clearly, while weight restoration remains the most proximal goal of treatment in offsetting the medical effects of starvation, it should not be expected that weight gain alone will ultimately confer commensurate psychological symptom remission. As precision medicine initiatives gain momentum, it is imperative that the core mechanisms underpinning psychological AN psychopathology are identified and examined. Future treatment development efforts ought to adopt a specific focus on the more rapid relief of psychological AN pathology, such that the mechanisms by which they drive behavioral symptoms may be targeted and ameliorated throughout treatment.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291718002088

References

Agras WS, Lock J, Brandt H, Bryson SW, Dodge E, Halmi KA, Jo B, Johnson C, Kaye W, Wilfley D and Woodside B (2014) Comparison of 2 family therapies for adolescent anorexia nervosa: a randomized parallel trial. *JAMA Psychiatry* 71, 1279–1286.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5.* Washington, DC: American Psychiatric Association.

Arcelus J, Mitchell AJ, Wales J and Nielson S (2011) Mortality rates in patients with anorexia nervosa and other eating disorders: a meta-analysis of 36 studies. *Archives of General Psychiatry* 68, 724–731.

Attia E, Haiman C, Walsh BT and Flater SR (1998) Does fluoxetine augment the inpatient treatment of anorexia nervosa? *American Journal of Psychiatry* 155, 548–551.

Attia E, Kaplan AS, Walsh BT, Gershkovich M, Yilmaz Z, Musante D and Wang Y (2011) Olanzapine versus placebo for out-patients with anorexia nervosa. *Psychological Medicine* 41, 2177–2182.

Ball J and Mitchell P (2004) A randomized controlled study of cognitive behavior therapy and behavioral family therapy for anorexia nervosa patients. *Eating Disorders* 12, 303–314.

Brambilla F, Garcia CS, Fassino S, Daga GA, Favaro A, Santonastaso P, Ramaciotti C, Bondi E, Mellado C, Borriello R and Montelone P (2007) Olanzapine therapy in anorexia nervosa: psychobiological effects. *International Clinical Psychopharmacology* 22, 197–204.

Court A, Mulder C, Kerr M, Yuen HP, Boasman M, Goldstone S, Fleming J, Weigall S, Derham H, Huang C, McGory P and Berger G (2010) Investigating the effectiveness, safety and tolerability of quetiapine in the treatment of anorexia nervosa in young people: a pilot study. *Journal of Psychiatric Research* 44, 1027–1034.

Cuijpers P, Weitz E, Cristea IA and Twisk J (2017) Pre-post effect sizes should be avoided in meta-analyses. *Epidemiology and Psychiatric Sciences* 26, 364–368.

Dalle-Grave R, Calugi S, Conti M, Doll H and Fairburn CG (2013) Inpatient cognitive behavior therapy for anorexia nervosa: a randomized controlled trial. *Psychotherapy and Psychosomatics* 82, 390–398.

DeJong H, Broadbent H and Schmidt U (2012) A systematic review of drop-out from treatment in outpatients with anorexia nervosa. *International Journal of Eating Disorders* 45, 635–647.

DuBois RH, Rogers RF, Franko DL, Eddy KT and Thomas JJ (2017) A network analysis investigation of the cognitive-behavioral theory of eating disorders. *Behaviour Research and Therapy* 97, 213–221.

Egger M, Smith GD, Schneider M and Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* 315, 629–634.

Eisler I, Dare C, Hodes M and Russell G (2000) Family therapy for adolescent anorexia nervosa: the results of a controlled comparison of two family interventions. *Journal of Child Psychology & Psychiatry* 41, 727–736.

Eisler I, Simic M, Hodsol J, Asen E, Berelowitz M, Connan F, Ellis G, Hugo P, Schmidt U, Treasure J, Yi I and Landau S (2016) A pragmatic randomized multi-centre trial of multifamily and single family therapy for adolescent anorexia nervosa. *BMC Psychiatry* 16, 422.

Fairburn CG (2005) Evidence-based treatment of anorexia nervosa. *International Journal of Eating Disorders* 37, S26–S30.

Fairburn CG, Shafran R and Cooper Z (1999) A cognitive behavioural theory of anorexia nervosa. *Behaviour, Research and Therapy* 37, 1–13.

Fassino S, Leombruni P, Abbate-Daga G, Brustolin A, Migliaretti G, Cavallo F and Rovera GG (2002) Efficacy of citalopram in anorexia nervosa: a pilot study. *European Neuropsychopharmacology* 12, 453–459.

Fichter MM, Quadflieg N, Crosby RD and Koch S (2017) Long-term outcome of anorexia nervosa: results from a large clinical longitudinal study. *International Journal of Eating Disorders* 50, 1018–1030.

Friedman K, Ramirez A, Murray SB, Anderson LK, Cuasck A, Boutelle KN and Kaye WH (2016) A narrative review of outcome studies for residential and partial hospital-based treatment of eating disorders. *European Eating Disorders Review* 24, 263–276.

Geist R, Heinmaa M, Stephens D, Davis R and Katzman DK (2000) Comparison of family therapy and family group psychoeducation in adolescents with anorexia nervosa. *Canadian Journal of Psychiatry* 45, 173–178.

Godart N, Berthoz S, Curt F, Perdereau F, Rein Z, Wallier J, Horreard AS, Kaganski I, Lucet R, Atger F, Corcos M, Fermanian J, Falissard B, Flament M, Eisler I and Jeameant P (2012) A randomized controlled trial of adjunctive family therapy and treatment as usual following inpatient treatment for anorexia nervosa adolescents. *PLoS ONE* 7, e28249.

Gowers SG, Clark A, Roberts C, Griffiths A, Edwards S, Bryan C, Smethurst N, Byford S and Barrett B (2007) Clinical effectiveness of treatments for anorexia nervosa in adolescents. *British Journal of Psychiatry* 191, 427–435.

Guarda AS, Schreyer CC, Fischer LK, Hansen JL, Coughlin JW, Kaminsky MJ, Attia E and Redgrave GW (2017) Intensive treatment for adults with anorexia nervosa: the cost of weight restoration. *International Journal of Eating Disorders* 50, 302–306.

Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, DeBeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P and Schünemann HJ (2011) GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology* 64, 383–394.

Hagman J, Gralla J, Sigel E, Ellert S, Dodge M, Gardner R, O'Lonergan T, Frank G and Wamboldt MZ (2011) A double-blind placebo controlled study of risperidone for the treatment of adolescents and young adults with anorexia nervosa: a pilot study. *Journal of the American Academy of Child & Adolescent Psychiatry* 50, 915–924.

Hedges LV, Tipton E and Johnson MC (2010) Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods* 1, 39–65.

Herpertz-Dahlmann B, Schwarte R, Krei M, Egberts K, Warnke A, Wewetzer C, Preiffer E, Fleischhaker C, Scherag A, Holtkamp K, Hagenah U, Bühren K, KLonrad K, Schmidt U, Schade-Brittinger C, Timmesfeld N and Dempfle A (2014) Day-patient treatment after short inpatient care versus continued inpatient treatment in adolescents with anorexia nervosa (ANDI): a multicentre, randomized, open-label non-inferiority trial. *The Lancet* 383, 5–11.

Herscovici CR, Kovalskys I and Orellana L (2017) An exploratory evaluation of the family meal intervention for adolescent anorexia nervosa. *Family Process* 56, 364–375.

Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Savocić J, Schulz KF, Weeks L and Sterne JAC (2011) The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal* 343.

Kaye WH, Nagata T, Weltzin TE, Hsu LKG, Sokol MS, McConaha C, Plotnicov KH, Weise J and Deep D (2001) Double-blind placebo-controlled administration of fluoxetine in restricting- and restricting-purging-type anorexia nervosa. *Biological Psychiatry* 49, 644–652.

Keel PK, Dorer DJ, Eddy KT, Franko D, Charatan DL and Herzog DB (2003) Predictors of mortality in eating disorders. *Archives of General Psychiatry* 60, 179–183.

Keel PK, Dorer DJ, Franko DL, Jackson SC and Herzog DB (2005) Postremission predictors of relapse in women with eating disorders. *American Journal of Psychiatry* 162, 2263–2268.

Le Grange D, Hughes EK, Court A, Yeo M, Crosby RD and Sawyer SM (2016) Randomized clinical trial of parent-focused treatment and family-based treatment for adolescent anorexia nervosa. *Journal of the American Academy of Child & Adolescent Psychiatry* 55, 683–692.

Lock J, Agras WS, Bryson S and Kraemer HC (2005) A comparison of short- and long-term family therapy for adolescent anorexia nervosa. *Journal of the American Academy of Child & Adolescent Psychiatry* **44**, 632–639.

Lock J, Le Grange D, Agras WS, Moye A, Bryson SW and Jo B (2010) Randomized clinical trial comparing family-based treatment with adolescent-focused individual therapy for adolescents with anorexia nervosa. *Archives of General Psychiatry* **67**, 1025–1032.

Madden S, Miscovic-Wheatley J, Wallis A, Kohn M, Lock J, Le Grange D, Jo B, Clarke S, Rhodes P, Hay P and Touyz SW (2015) A randomized controlled trial of in-patient treatment for anorexia nervosa in medically unstable adolescents. *Psychological Medicine* **45**, 415–427.

Matthews RRS, Hall WD, Vos T, Patton GC and Degenhardt L (2011) What are the major drivers of prevalent disability burden in young Australians? *Medical Journal of Australia* **194**, 232–235.

McIntosh VVW, Jordan J, Carter FA, Luty SE, McKenzie JM, Bulik CM, Frampton CMA and Joyce PR (2005) Three psychotherapies for anorexia nervosa: a randomized, controlled trial. *American Journal of Psychiatry* **162**, 741–747.

Mitchell JE and Crow S (2006) Medical complications of anorexia nervosa and bulimia nervosa. *Current Opinion in Psychiatry* **19**, 438–443.

Moher D, Liberati A, Tetzlaff J and Altman DG, The PRISMA Group. (2009). PRISMA group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine* **6**: e1000097.

Mondraty N, Birmingham CL, Touyz SW, Sunkadov V, Chapman L and Beumont P (2005) Randomized controlled trial of olanzapine in the treatment of cognitions in anorexia nervosa. *Australasian Psychiatry* **13**, 72–75.

Parling T, Cernvall M, Ramklint M, Holmgren S and Ghaderi A (2016) A randomised trial of acceptance and commitment therapy for anorexia nervosa after daycare treatment, including five-year follow-up. *BMC Psychiatry* **272**.

Peters JL, Sutton AJ, Jones DR, Abrams KR and Rushton L (2008) Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology* **61**, 991–996.

Powers PS, Klabunde M and Kaye WH (2012) Double-blind placebo-controlled trial of quetiapine in anorexia nervosa. *European Eating Disorders Review* **20**, 331–334.

Rigaud D, Brondel L, Poupard AT, Talonneau I and Brun JM (2007) A randomized trial on the efficacy of a 2-month tube feeding regimen in anorexia nervosa: a 1-year follow-up study. *Clinical Nutrition* **26**, 421–429.

Robin AL, Siegel PT, Moye AW, Gilroy M, Dennis AB and Sikand A (1999) A controlled comparison of family versus individual therapy for adolescents with anorexia nervosa. *Journal of the American Academy of Child & Adolescent Psychiatry* **38**, 1482–1489.

Rücker G, Cates CJ and Schwarzer G (2017) Methods for including information from multi-arm trials in pairwise meta-analysis. *Research Synthesis Methods* **8**, 392–403.

Russell J, Maguire S, Hunt GE, Kesby A, Suraev A, Stuart J, Booth J and McGregor IS (2018) Intranasal oxytocin in the treatment of anorexia nervosa: randomized controlled trial during re-feeding. *Psychoneuroendocrinology* **87**, 83–92.

Schmidt U, Oldershaw A, Jicji F, Sternheim L, Startup H, McIntosh V, Jordan J, Tchanturia K, Wolff G, Rooney M, Landau S and Treasure J (2012) Out-patient psychological therapies for adults with anorexia nervosa: randomized controlled trial. *British Journal of Psychiatry* **201**, 392–399.

Schmidt U, Magill N, Renwick B, Keyes A, Kenyon M, Dejong H, Lose A, Broadbent H, Loomes R, Yasin H, Watson C, Ghelani S, Bonin EM, Serpell L, Richards L, Johnson-Sabine E, Boughton N, Whitehead L, Beecham J, Treasure J and Landau S (2015) The Maudsley outpatient study of treatments for anorexia nervosa and related conditions (MOSAIC): comparison of the Maudsley model of anorexia nervosa treatment for adults (MANTRA) with specialist supportive clinical management (SSCM) in outpatients with broadly defined anorexia nervosa: a randomized controlled trial. *Journal of Consulting and Clinical Psychology* **83**, 796–807.

Smith C, Fogarty S, Touyz SW, Madden S, Buckett G and Hay P (2014) Acupuncture and acupressure and massage health outcomes for patients with anorexia nervosa: findings from a pilot randomized controlled trial and patient interviews. *Journal of Alternative and Complementary Medicine* **20**, 103–112.

Steinhaussen HC (2002) The outcome of anorexia nervosa in the 20th century. *American Journal of Psychiatry* **159**, 1284–1293.

Touyz SW, Le Grange D, Lacey H, Hay P, Smoth R, Maguire S, Bamford B, Pike KM and Crosby RD (2013) Treating severe and enduring anorexia nervosa: a randomized controlled trial. *Psychological Medicine* **43**, 2501–2511.

Walsh BT, Kaplan AS, Attia E, Olmsted M, Parides M, Carter JC, Pike KM, Devlin MJ, Woodside B, Roberto CA and Rockert W (2006) Fluoxetine after weight restoration in anorexia nervosa: a randomized controlled trial. *JAMA* **295**, 2605–2612.

Watson HJ and Bulik CM (2013) Update on the treatment of anorexia nervosa: review of clinical trials, practice guidelines and emerging interventions. *Psychological Medicine* **43**, 2477–2500.

Whitney J, Murphy T, Landau S, Gavan K, Todd G, Whitaker W and Treasure J (2011) A practical comparison of two types of family intervention: an exploratory RCT of family day workshops and individual family work as a supplement to inpatient care for adults with anorexia nervosa. *European Eating Disorders Review* **20**, 142–150.

Woodside DB, Carter JC and Blackmore E (2004) Predictors of premature termination of inpatients treatment for anorexia nervosa. *American Journal of Psychiatry* **161**, 2277–2281.

Zipfel S, Wild B, Grob B, Friederich HC, Teufel M, Schellberg D, Giel KE, de Zwaan M, Dinkel A, Herpertz S, Burgmer M, Löwe B, Tagay S, von Wietersheim J, Zeeck A, Schade-Brittinger C, Schauenburg H and Herzog W (2014) Focal dynamic psychotherapy, cognitive behavior therapy, and optimized treatment as usual in outpatients with anorexia nervosa (ANTOP study): randomized controlled trial. *The Lancet* **383**, 127–137.