Supplementary Materials for "A comprehensive updated cross-sectional and longitudinal meta-analysis of cytokines in eating disorders"

Table of Contents

1. **Quality Assessment: Newcastle-Ottawa Scale adapted for cross-sectional studies**

Studies are allocated stars (up to a total of 11) based on meeting certain criteria, described below. The greater number of stars allocated indicates a higher quality.

Selection: (Maximum 5 stars)

1) Representativeness of the sample:

a) Truly representative of the average in the target population (all subjects or random sampling) **

b) Somewhat representative of the average in the target population (non-random sampling) *

c) Selected group of users or no description of the sampling strategy.

2) Sample size:

a) Justified and satisfactory *

b) Not justified

3) Ascertainment of the exposure (disease i.e. ED):

a) Validated measurement tool in both ED and HC group **

b) Non-validated measurement tool, but the tool is available or described in both ED and HC group or validated measurement tool described in one group (ED or HC) *

c) No description of the measurement tool or non-validated measurement tool, but the tool is available or described for only one group (ED or HC)

Comparability: (Maximum 3 stars)

1) The subjects in different outcome groups are comparable, based on the study design. Confounding factors are controlled through pre-analysis techniques such as sample matching, data cleaning or eligibility criteria.

a) The study controls for the most important factor (age) *

b) The study controls for an additional important factor (sex) *

c) The study controls for an additional important factor (e.g., smoking status, presence of inflammatory condition, use of medication) *

Outcome: (Maximum 3 stars)

1) Measurement of outcome:

a) Validated measurement method (interassay CV included) **

b) Non-validated measurement method, but the method is available or described *

c) No description of the measurement tool

2) Statistical test:

a) The statistical test used to analyse the data is clearly described and appropriate, and the measurement of the association is presented, including confidence intervals and the probability level (p value). *

b) The statistical test is not appropriate, not described or incomplete.

This scale was adapted from the Newcastle-Ottawa Quality Assessment Scale for cohort studies to perform a quality assessment of cross-sectional studies for use in the following systematic review: Herzog R, Álvarez-Pasquin MJ, Díaz C, Del Barrio JL, Estrada JM, Gil Á, 2013. Are healthcare workers' intentions to vaccinate related to their knowledge, beliefs and attitudes? A systematic review. BMC Public Health. 13, 154. We further adapted the scale from ascertainment of outcome (e.g. independent blind assessment/record linkage/self-report) to measurement of outcome to account for the methods used in this field of research.

1.1. Summary of quality assessment

About half of the studies ensured representativeness of their sample (n=21), and few used a validated measurement tool in both the ED and HC groups to ascertain diagnosis (n=16). Some studies either used a non-validated measurement tool or only used the tool in the ED group (n=12). Fifteen studies had no description of the measurement tool (n=14). Three studies provided justification for their sample size (n=3). In terms of comparability, most studies accounted for age as a confounding variable (n=26), and almost all studies accounted for an additional confounding factor (e.g. BMI, sex; smoking status; n=41). Few studies explicitly reported using a validated method to measure the inflammatory factor level, including the interassay CV (n=19), 23 studies used a non-validated measurement method that was described and one study did not describe their measurement tool. All studies presented the measurement of the association with CIs and the probability level (*p*-values) (n=43).

Table S1. Quality assessment of studies included in the systematic review and meta-analysis using the Newcastle-Ottowa Scale adapted for cross-sectional studies

	Selection (S)			Comparability (C)			Outcome (O)		Subtotal Assessment			
	Representativeness of the sample	Sample size	Ascertainment of the exposure	Most important confound controlled for	Additional important confound controlled for 1	Additional important confound controlled for 2	Measurement of outcome	Statistical test	S[a]	C[b]	O[c]	Conclusion
Agnello et al. 2012	*			*	*	*	*	*	Poor	Good	Good	Fair
Ahren-Moonga et al. 2011	*			*	*	*	*	*	Poor	Good	Good	Fair
Allende et al. 1998			**	*	*	*	*	*	Fair	Good	Good	Good
Amerio et al. 2023				*	*	*	**	*	Poor	Good	Good	Fair
Brambilla et al. 1998			*	*	*	*	**	*	Poor	Good	Good	Fair
Brambilla et al. 2001			**	*	*	*	**	*	Fair	Good	Good	Good
Breithaupt et al. 2024	*		*	*	*	*	*	*	Fair	Good	Good	Good
Caldas et al. 2022	*		**	*	*	*	*	*	Good	Good	Good	Good
Campanile et al. 2024			**		*	*	**	*	Fair	Good	Good	Good
Caroleo et al. 2019	*		**			*	**	*	Good	Fair	Good	Good
Caso et al. 2020	*		*	*	*	*	*	*	Fair	Good	Good	Good
Caso et al. 2022			*	*	*	*	*	*	Poor	Good	Good	Fair
Corcos et al. 2001			*		*	*		*	Poor	Good	Fair	Poor
Dalton et al. 2018	*		**		*	*	*	*	Good	Good	Good	Good
Dalton et al. 2020	*		**		*	*	*	*	Good	Good	Good	Good
Diaz-Marsa et al. 2021	*		**		*	*	*	*	Good	Good	Good	Good
Diaz-Marsa et al. 2024	**		*				*	*	Good	Poor	Good	Fair
Di Paolo et al. 2024	*		*		*	*	*	*	Fair	Good	Good	Good
Dolezalova et al. 2007			*	*	*	*	**	*	Poor	Good	Good	Fair
Elegido et al. 2019	**	*	**	*	*		*	*	Good	Good	Good	Good

Gabriel et al. 2023	*		*				*	*	Fair	Poor	Good	Poor
Germain et al. 2016					*	*	**	*	Poor	Good	Good	Fair
Jiskra et al. 2000					*	*	*	*	Poor	Good	Good	Fair
Karczewska-Kupczewska et al. 2012		*			*	*	**	*	Poor	Good	Good	Fair
Karczewska-Kupczewska et al. 2013	*		*		*	*	**	*	Fair	Good	Good	Good
Käver et al. 2024	**		**	*	*	*	*	*	Good	Good	Good	Good
Keeler et al. 2021	**	*	**		*	*	*	*	Good	Good	Good	Good
Keeler et al. 2024	**		**	*	*	*	*	*	Good	Good	Good	Good
Krizova et al. 2002				*	*		*	*	Poor	Good	Good	Fair
Krizova et al. 2008				*	*		*	*	Poor	Good	Good	Fair
Misra et al. 2006				*	*		**	*	Poor	Good	Good	Fair
Monteleone et al. 1999			**		*	*	**	*	Fair	Good	Good	Good
Nagata et al. 2006			**	*	*	*	**	*	Fair	Good	Good	Good
Nakai et al. 1999	*			*	*	*	**	*	Poor	Good	Good	Fair
Nakai et al. 2000	*			*	*	*	**	*	Poor	Good	Good	Fair
Nilsson et al. 2020	**		**	*	*		*	*	Good	Good	Good	Good
Nogueira et al. 2010			*		*		*	*	Poor	Fair	Good	Poor
Ostrowska et al. 2015			*		*	*	**	*	Poor	Good	Good	Fair
Ostrowska et al. 2016			*	*	*		**	*	Poor	Good	Good	Fair
Pomeroy et al. 1994					*	*	*	*	Poor	Good	Good	Fair
Roczniak et al. 2020	*		*		*	*	**	*	Fair	Good	Good	Good
Roubalova et al. 2021	*		*		*	*	*	*	Fair	Good	Good	Good
Ruiz-Guerrero et al. 2022	**		**	*	*		*	*	Good	Good	Good	Good
Ruiz-Guerrero et al. 2023	*		**		*	*	*	*	Good	Good	Good	Good
Shimizu et al. 2005				*	*	*	*	*	Poor	Good	Good	Fair
Solis et al., 2002	*		**	*	*		*		Good	Good	Fair	Good

Tabasi et al. 2020		*	*	*	*	**	*	Poor	Good	Good	Fair
Terra et al. 2013		**		*	*	**	*	Fair	Good	Good	Good
Tyszkiewicz-Nwafor et al. 2022	*	**		*	*	**	*	Good	Good	Good	Good
Vaisman et al., 1996			*	*		*	*	Poor	Good	Good	Fair
Vaisman et al. 2004			*	*		**	*	Poor	Good	Good	Fair
Victor et al. 2015		*	*	*	*	**	*	Poor	Good	Good	Fair
Yasuhara et al. 2007	*		*	*		*	*	Poor	Good	Good	Fair

[a]Domain scored: 0-1 (Poor), 2 (Fair), 3+ (Good); [b]Domain scored: 0 (Poor), 1 (Fair), 2+ (Good); [c]Domain scored: 0 (Poor); 1 (Fair); 2+ (Good).

2. Cumulative meta-analysis figures

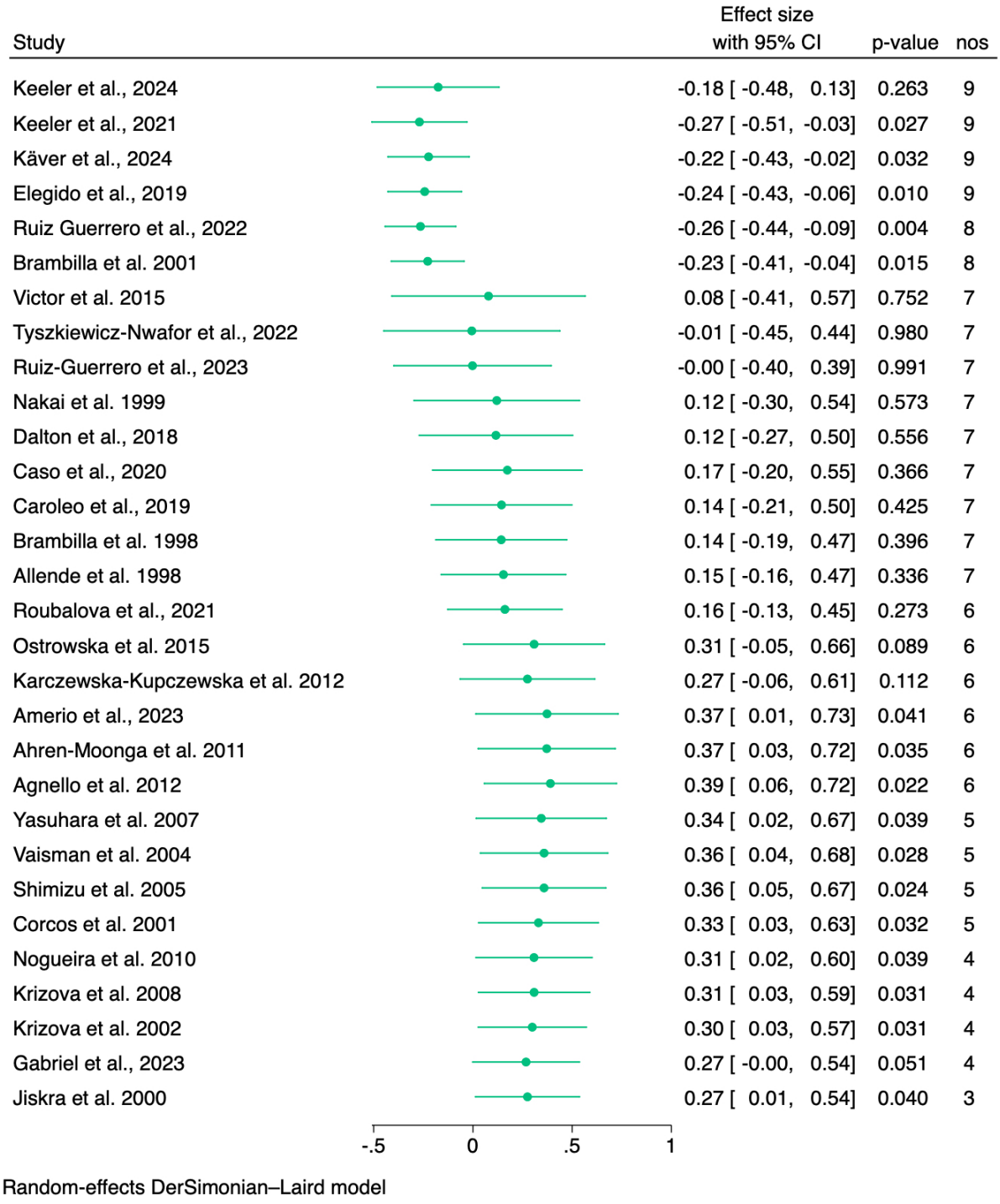

Figure S1. Cumulative forest plot for meta-analysis of tumour necrosis factor-α (TNF-α) concentrations between anorexia nervosa and healthy controls, descending according to Newcastle-Ottawa (NOS) quality assessment score.

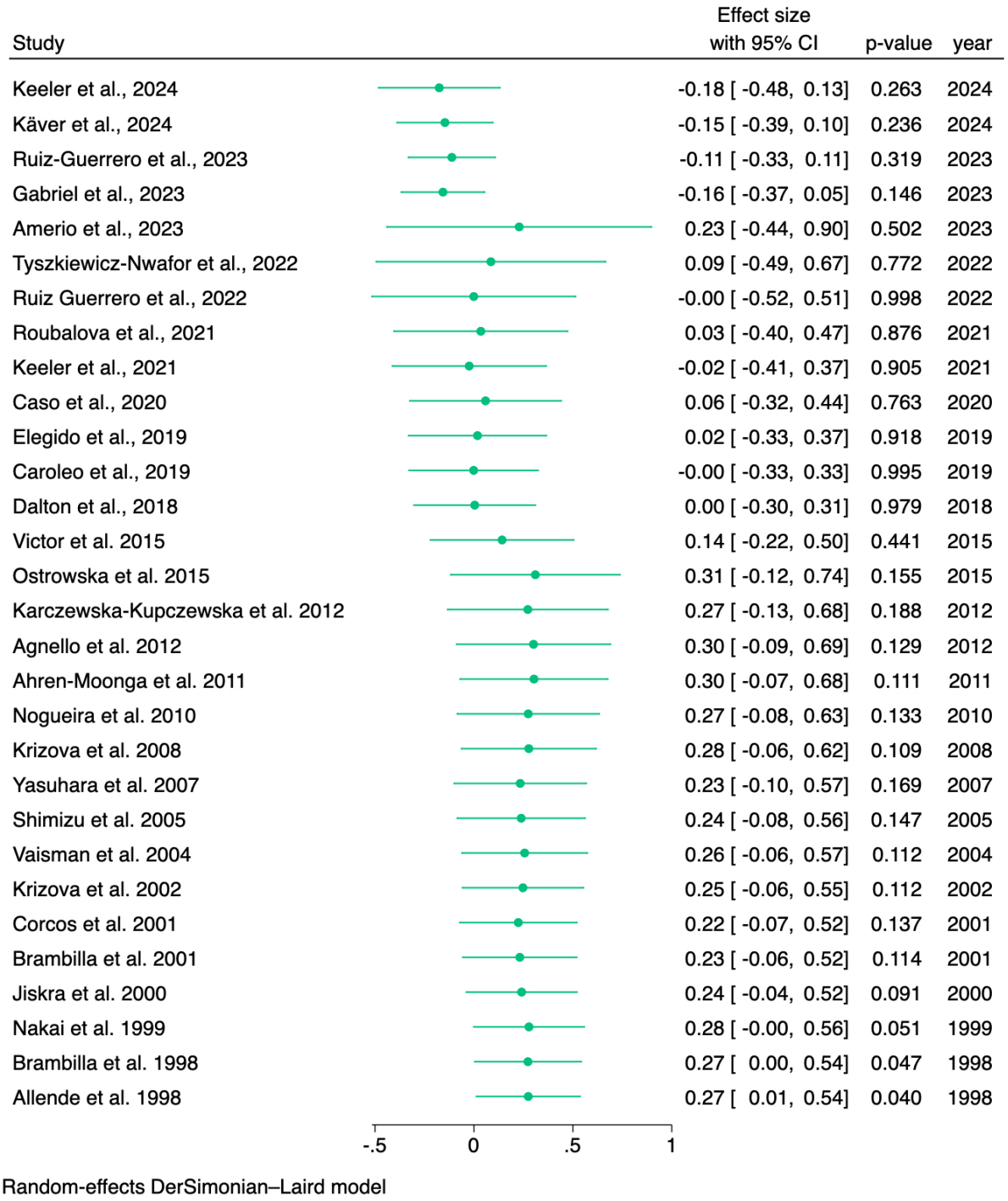

Figure S2. Cumulative forest plot for meta-analysis of tumour necrosis factor-α (TNF-α) concentrations between anorexia nervosa and healthy controls, descending according to year of publication.

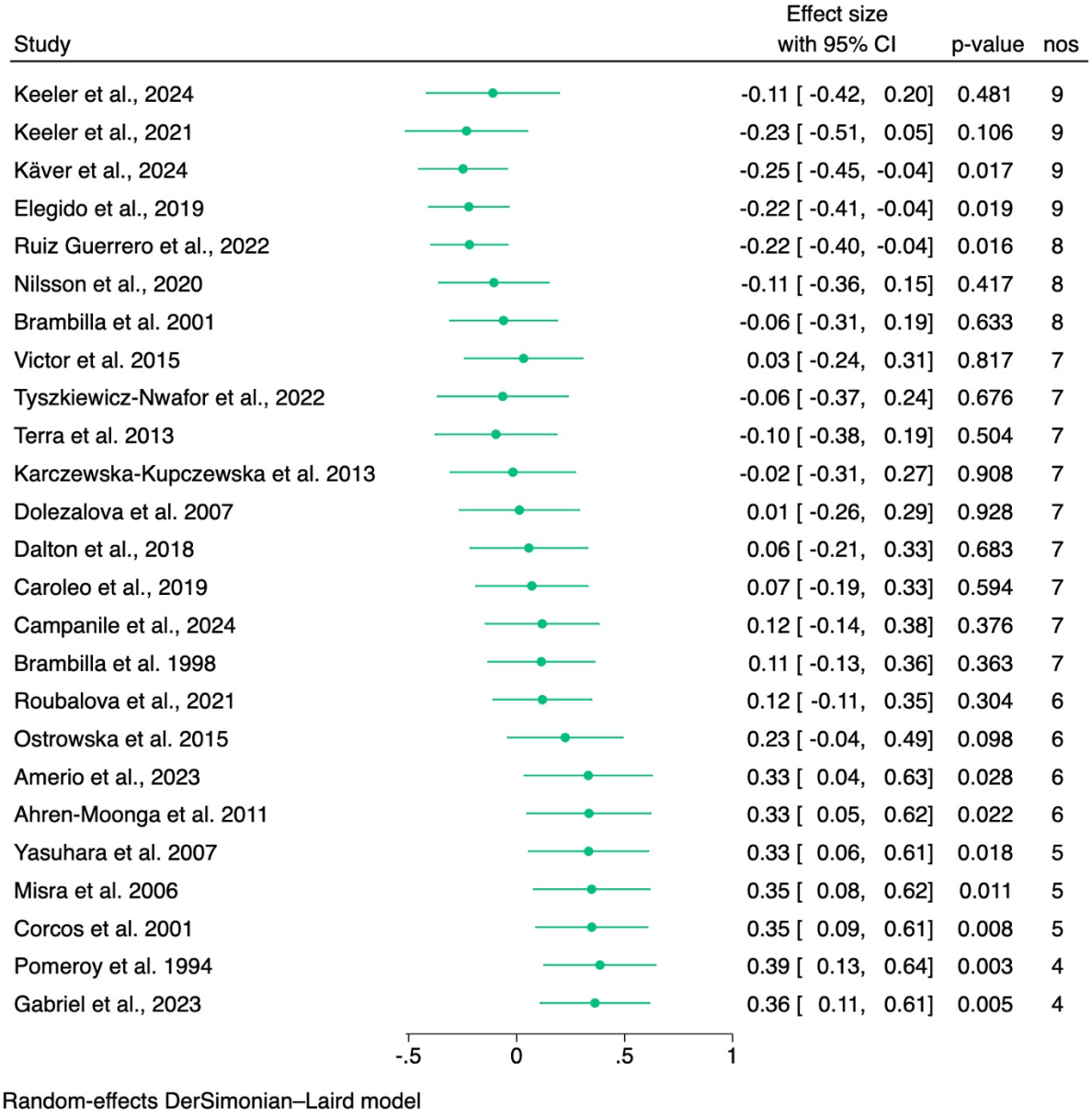

Figure S3. Cumulative forest plot for meta-analysis of interleukin-6 (IL-6) concentrations between anorexia nervosa and healthy controls, descending according to Newcastle-Ottawa (NOS) quality assessment score.

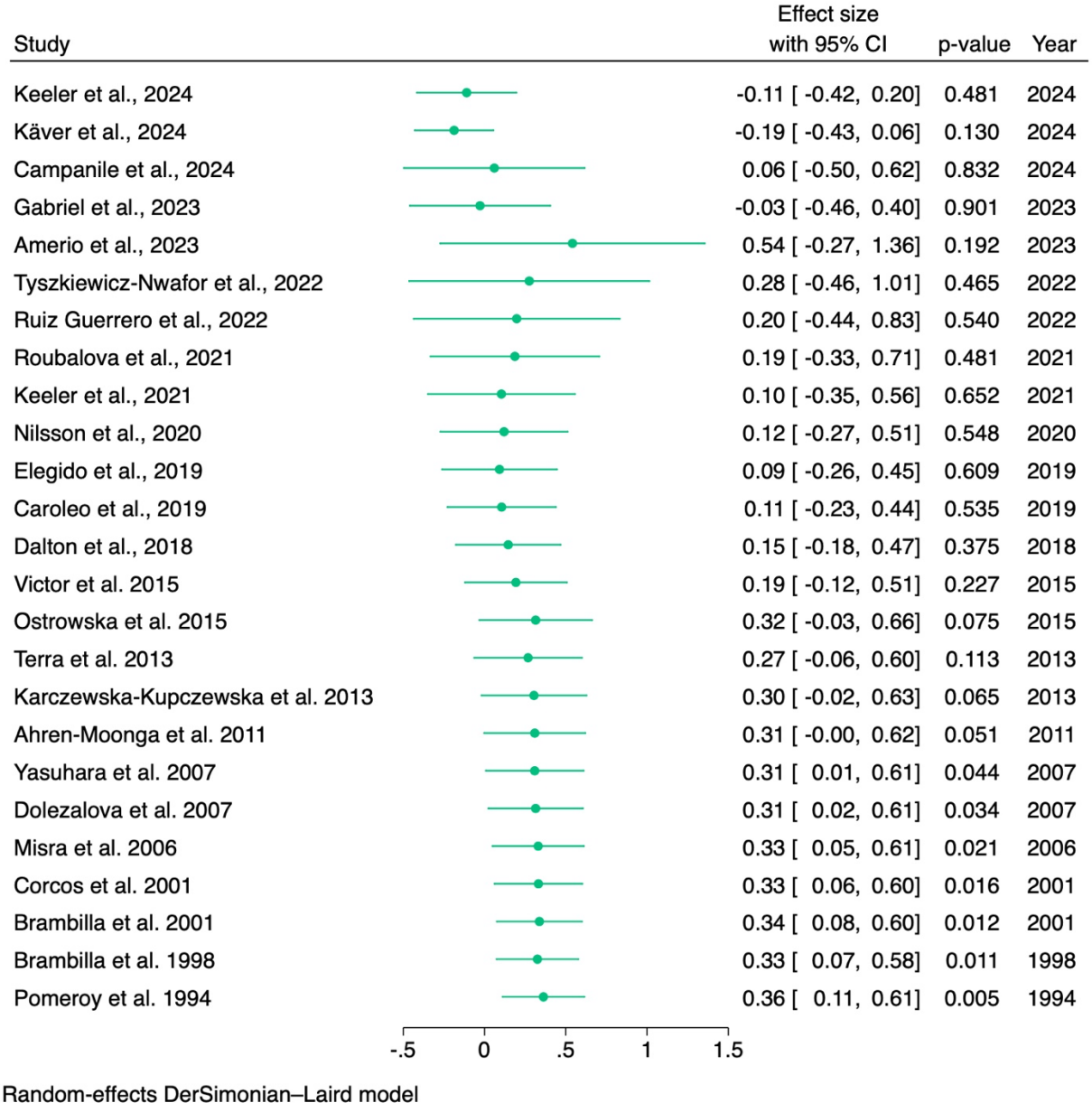

Figure S4. Cumulative forest plot for meta-analysis of interleukin-6 (IL-6) concentrations between anorexia nervosa and healthy controls, descending according to year of publication.

3. Funnel plots

3.1. Anorexia nervosa cross-sectional analyses

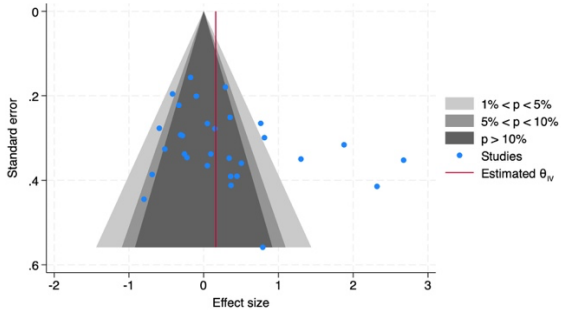

Figure S5. Funnel plot for cross-sectional meta-analysis of tumour necrosis factor-α (TNF-α) concentrations between anorexia nervosa and healthy controls

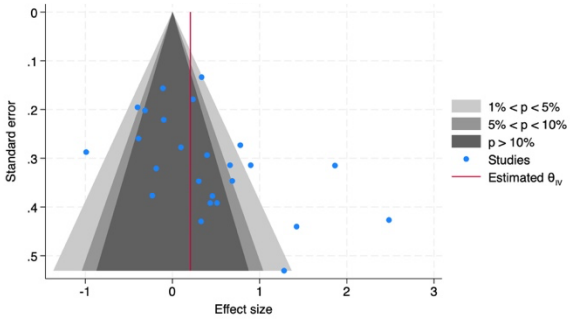

Figure S6. Funnel plot for cross-sectional meta-analysis of interleukin-6 (IL-6) concentrations between anorexia nervosa and healthy controls

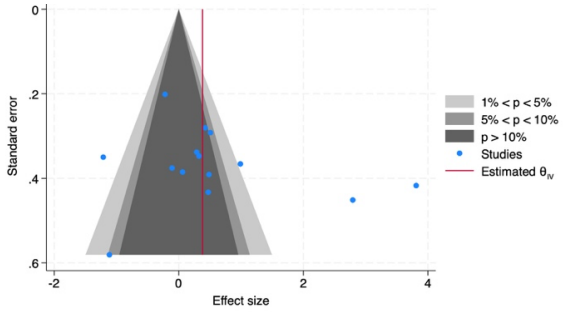

Figure S7. Funnel plot for cross-sectional meta-analysis of interleukin-1β (IL-1β) concentrations between anorexia nervosa and healthy controls

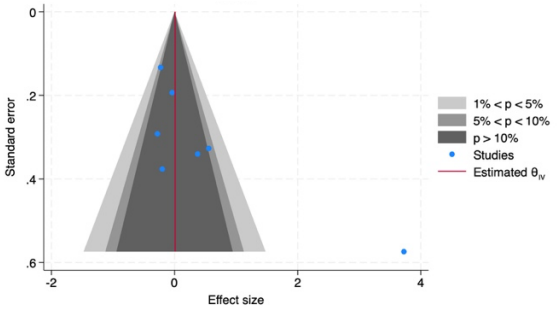

Figure S8. Funnel plot for cross-sectional meta-analysis of interleukin-10 (IL-10) concentrations between anorexia nervosa and healthy controls

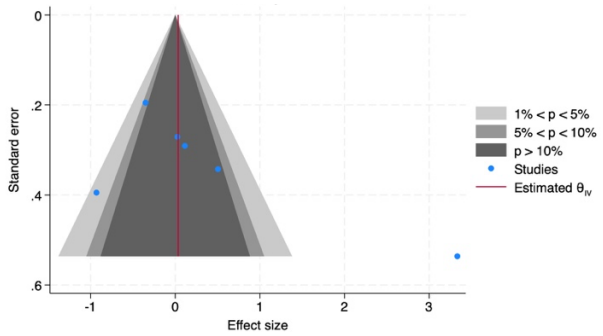

Figure S9. Funnel plot for cross-sectional meta-analysis of interferon-γ (IFN-γ) concentrations between anorexia nervosa and healthy controls

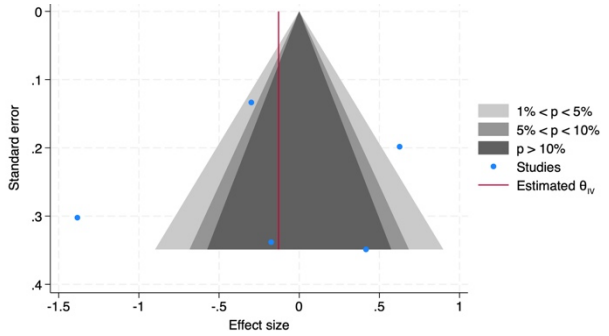

Figure S10. Funnel plot for cross-sectional meta-analysis of interleukin-8 (IL-8) concentrations between anorexia nervosa and healthy controls

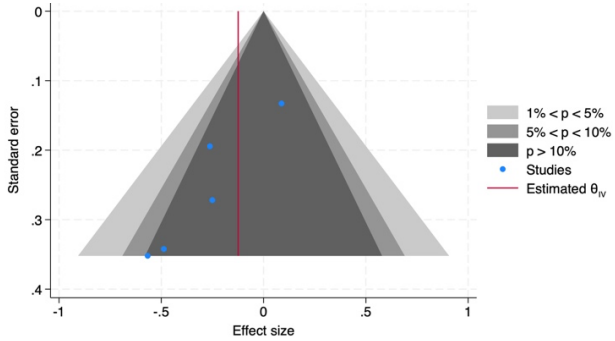

Figure S11. Funnel plot for cross-sectional meta-analysis of monocyte chemoattractant protein-1 (MCP-1) concentrations between anorexia nervosa and healthy controls

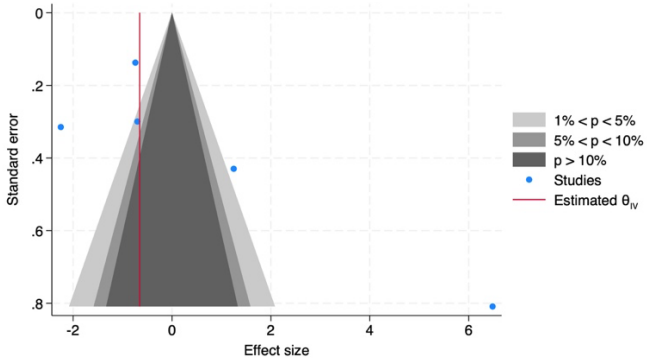

Figure S12. Funnel plot for cross-sectional meta-analysis of transforming growth factor-β (TGF-β) concentrations between anorexia nervosa and healthy controls

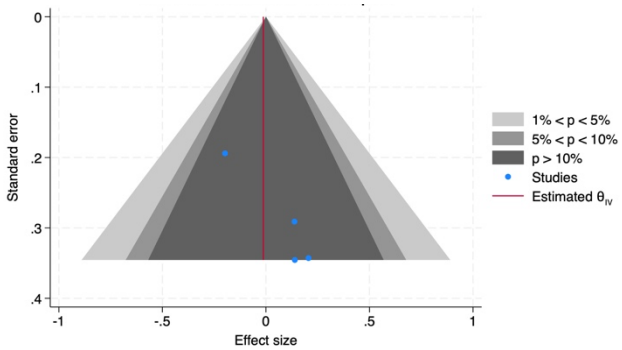

Figure S13. Funnel plot for cross-sectional meta-analysis of interleukin-4 (IL-4) concentrations between anorexia nervosa and healthy controls

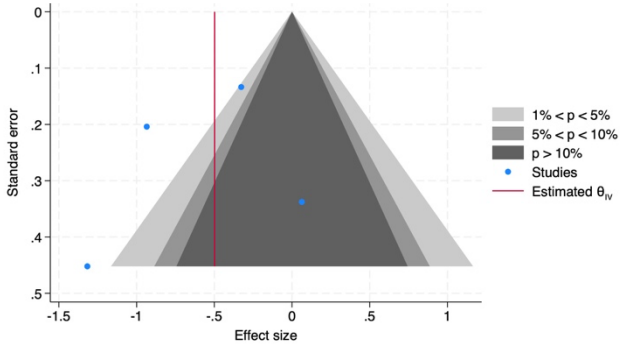

Figure S14. Funnel plot for cross-sectional meta-analysis of interleukin-7 (IL-7) concentrations between anorexia nervosa and healthy controls

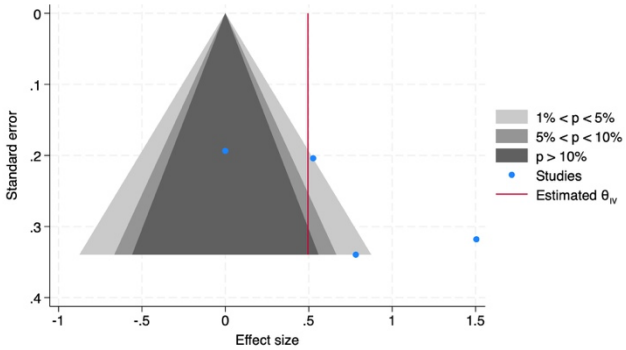

Figure S15. Funnel plot for cross-sectional meta-analysis of interleukin-15 (IL-15) concentrations between anorexia nervosa and healthy controls

3.2. Bulimia nervosa cross-sectional analyses

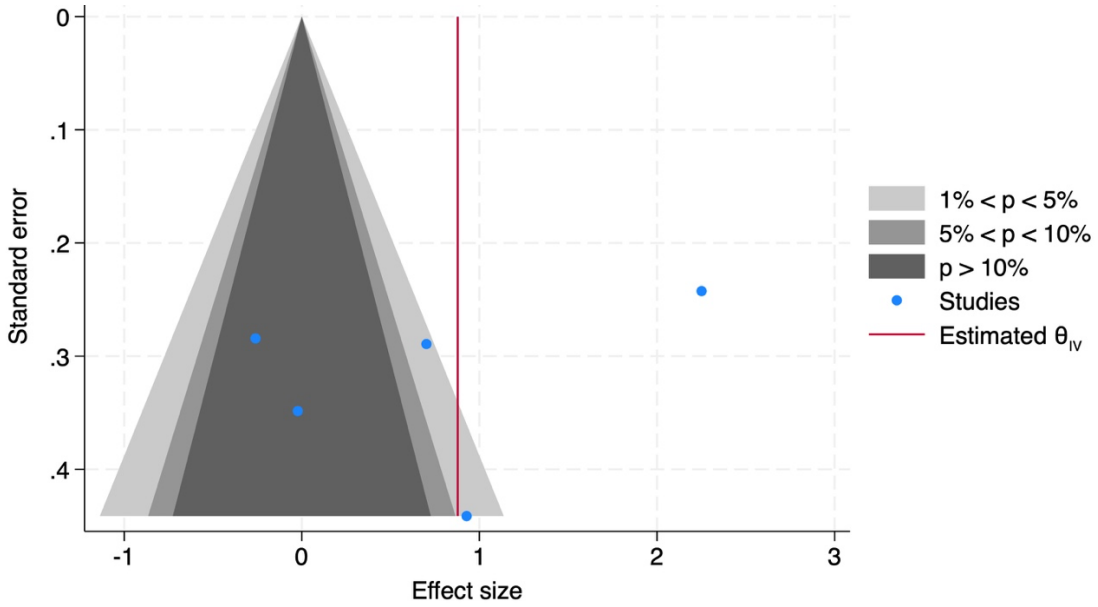

Figure S16. Funnel plot for cross-sectional meta-analysis of interleukin-6 (IL-6) concentrations between bulimia nervosa and healthy controls

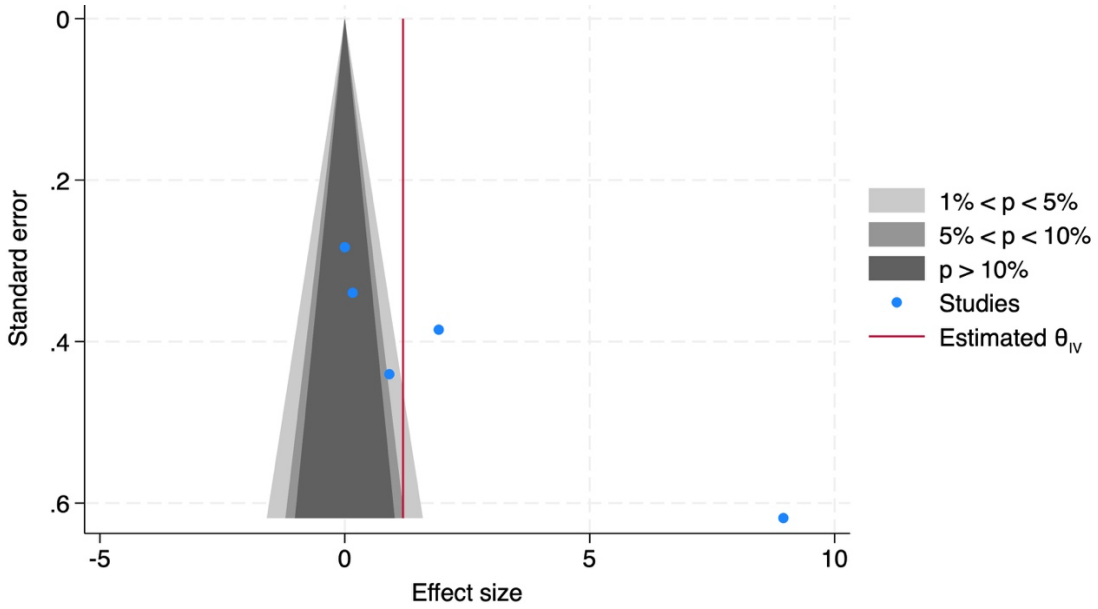

Figure S17. Funnel plot for cross-sectional meta-analysis of tumour necrosis factor-α (TNF-α) concentrations between bulimia nervosa and healthy controls

3.3. Anorexia nervosa longitudinal analyses

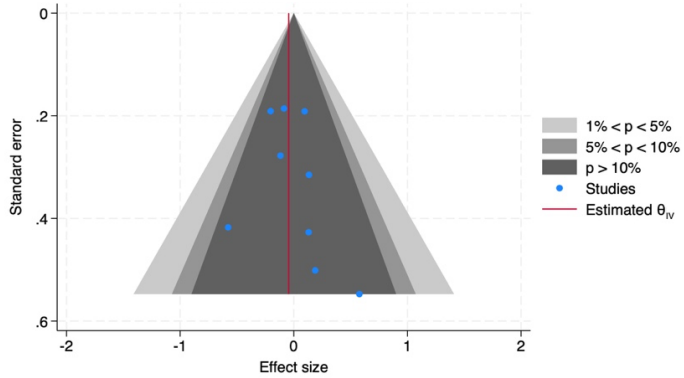

Figure S18. Funnel plot for meta-analysis of tumour necrosis factor-α (TNF-α) concentrations between anorexia nervosa at baseline and follow-up

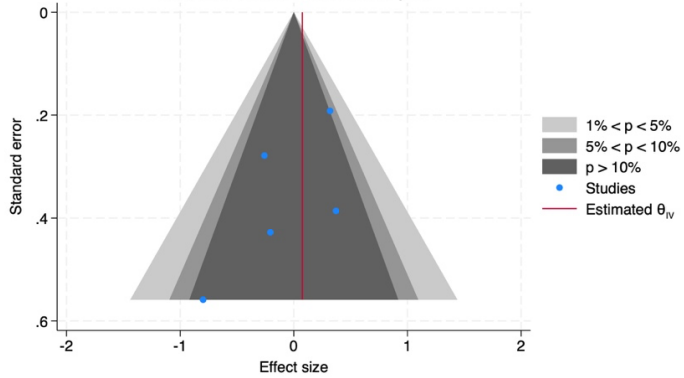

Figure S19. Funnel plot for meta-analysis of interleukin-1β (IL-β) concentrations between anorexia nervosa at baseline and follow-up

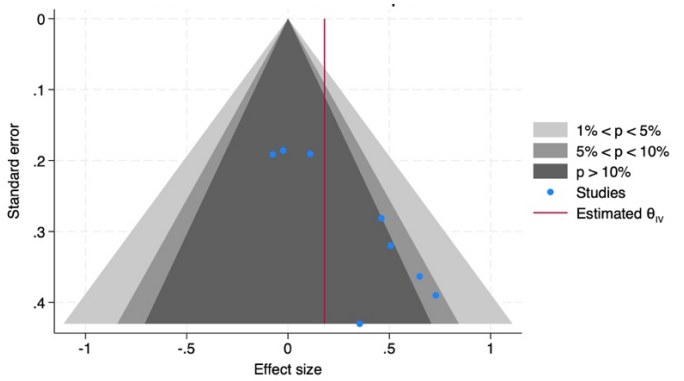

Figure S20. Funnel plot for meta-analysis of interleukin-6 (IL-6) concentrations between anorexia nervosa at baseline and follow-up

4. Trim and fill figures

4.1. Trim and fill funnel plots for anorexia nervosa cross-sectional analyses

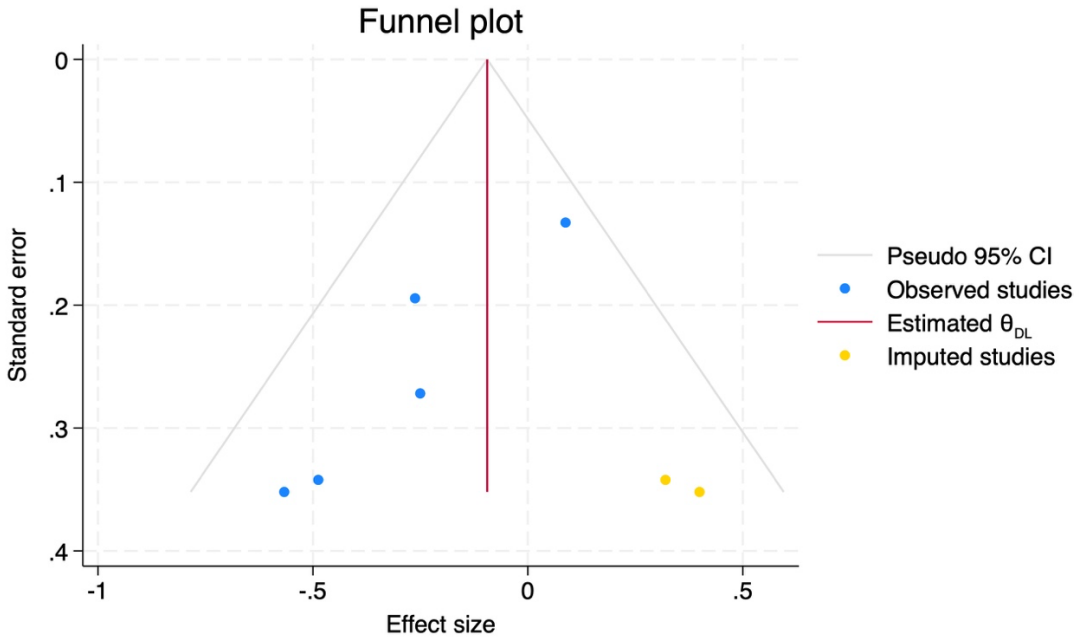

Figure S21. Trim-fill funnel plot for cross-sectional meta-analysis of monocyte chemoattractant protein-1 (MCP-1) concentrations between anorexia nervosa and healthy controls

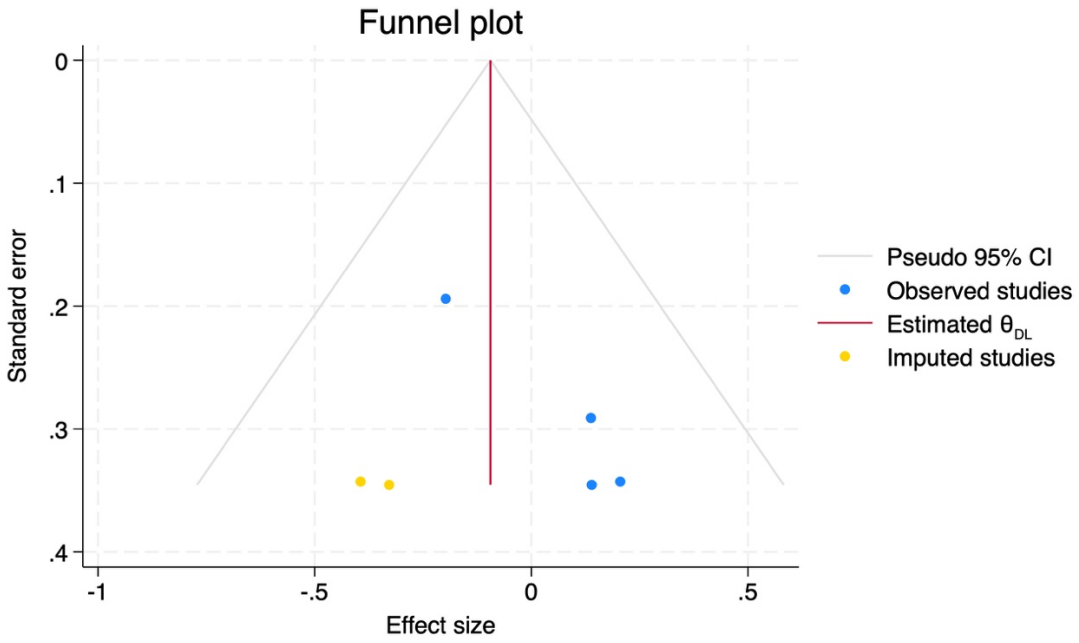

Figure S22. Trim-fill funnel plot for cross-sectional meta-analysis of interleukin-4 (IL-4) concentrations between anorexia nervosa and healthy controls

4.2. Trim and fill funnel plots for anorexia nervosa longitudinal analyses

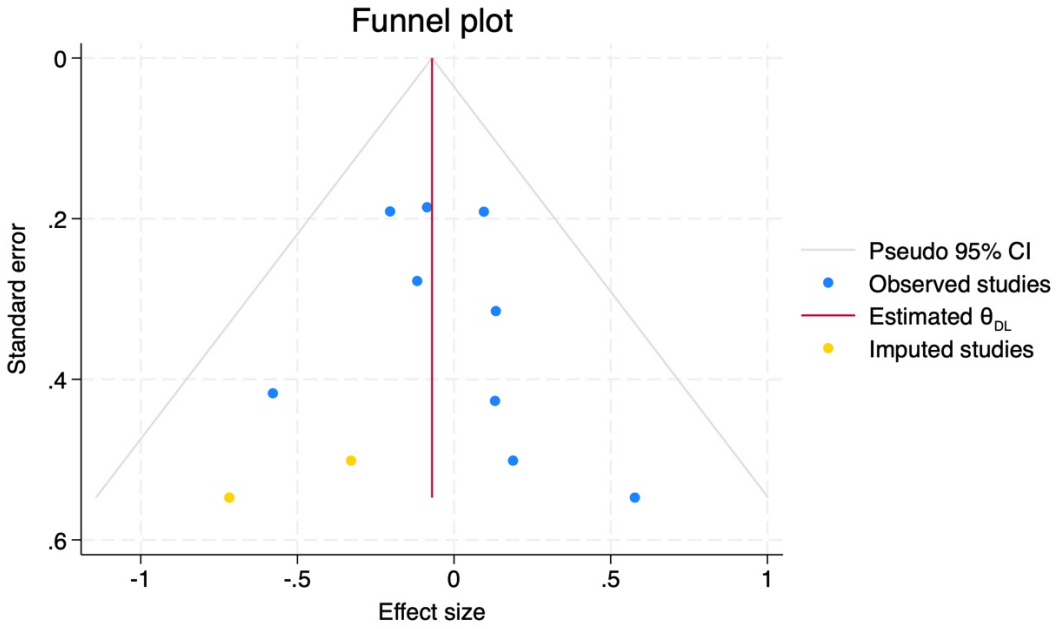

Figure S23. Trim-fill funnel plot for meta-analysis of tumour necrosis factor-α (TNF-α) concentrations between anorexia nervosa at baseline and follow-up

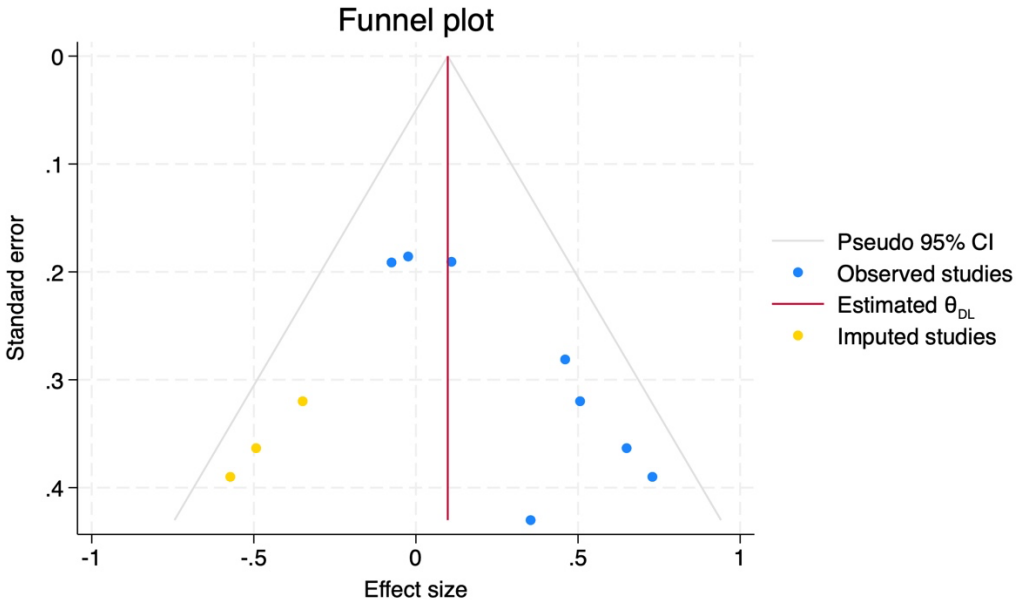

Figure S24. Trim-fill funnel plot for meta-analysis of interleukin-6 (IL-6) concentrations between anorexia nervosa at baseline and follow-up

5. Forest plots

5.1. Subgroup cross-sectional meta-analyses between anorexia nervosa and healthy controls according to anorexia nervosa subtype

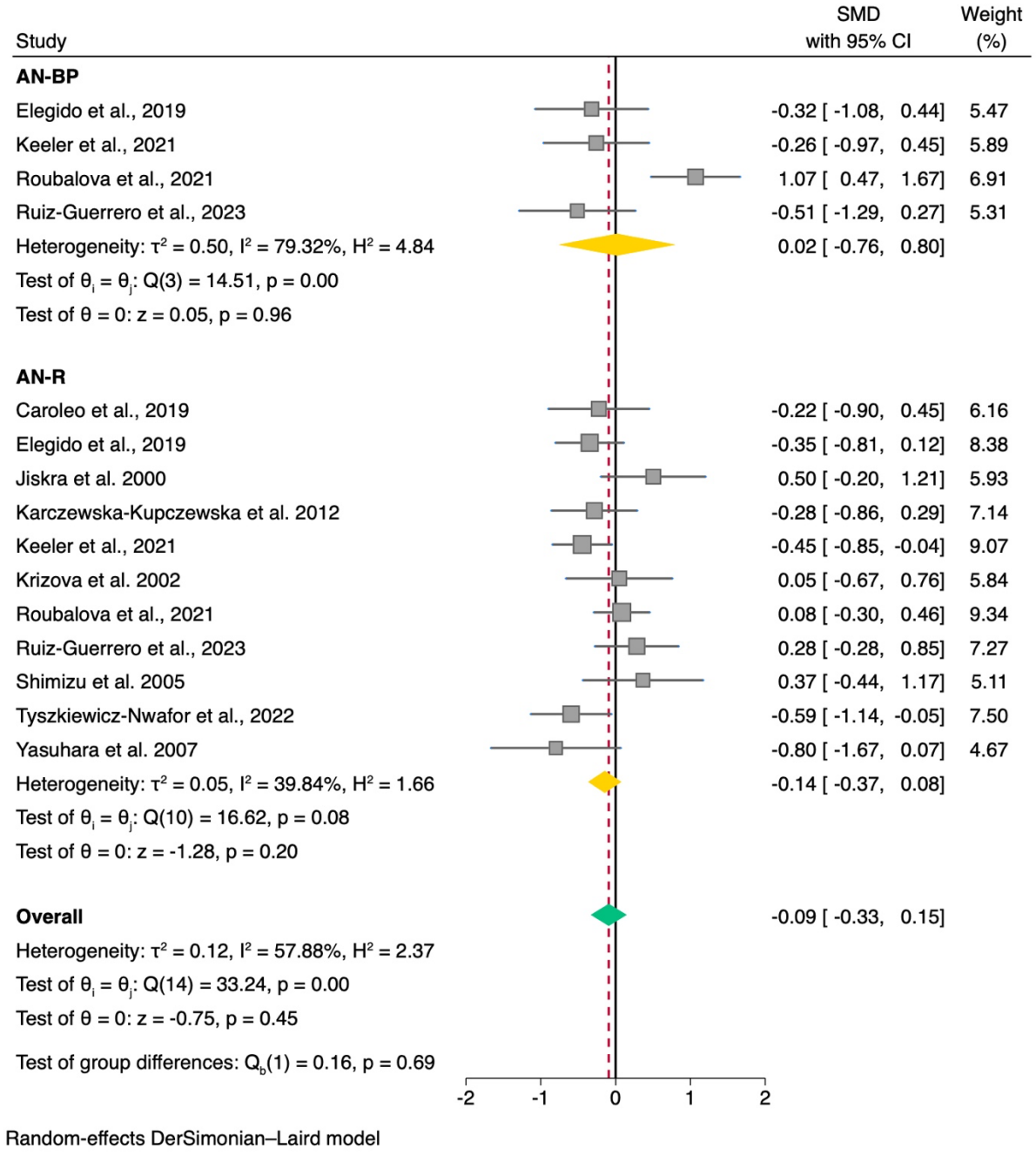

Figure S25. Forest plot of standardized mean difference in tumour necrosis factor-α (TNF-α) between anorexia nervosa (AN) participants and healthy controls stratified by AN subtype. Zero is the line of no effect, and points to the right of zero indicate an elevation of the cytokine in AN compared to HCs. Abbreviations: AN-BP = AN binge-purge subtype; AN-R = AN restricting subtype; CI = confidence intervals; SMD = standardized mean difference.

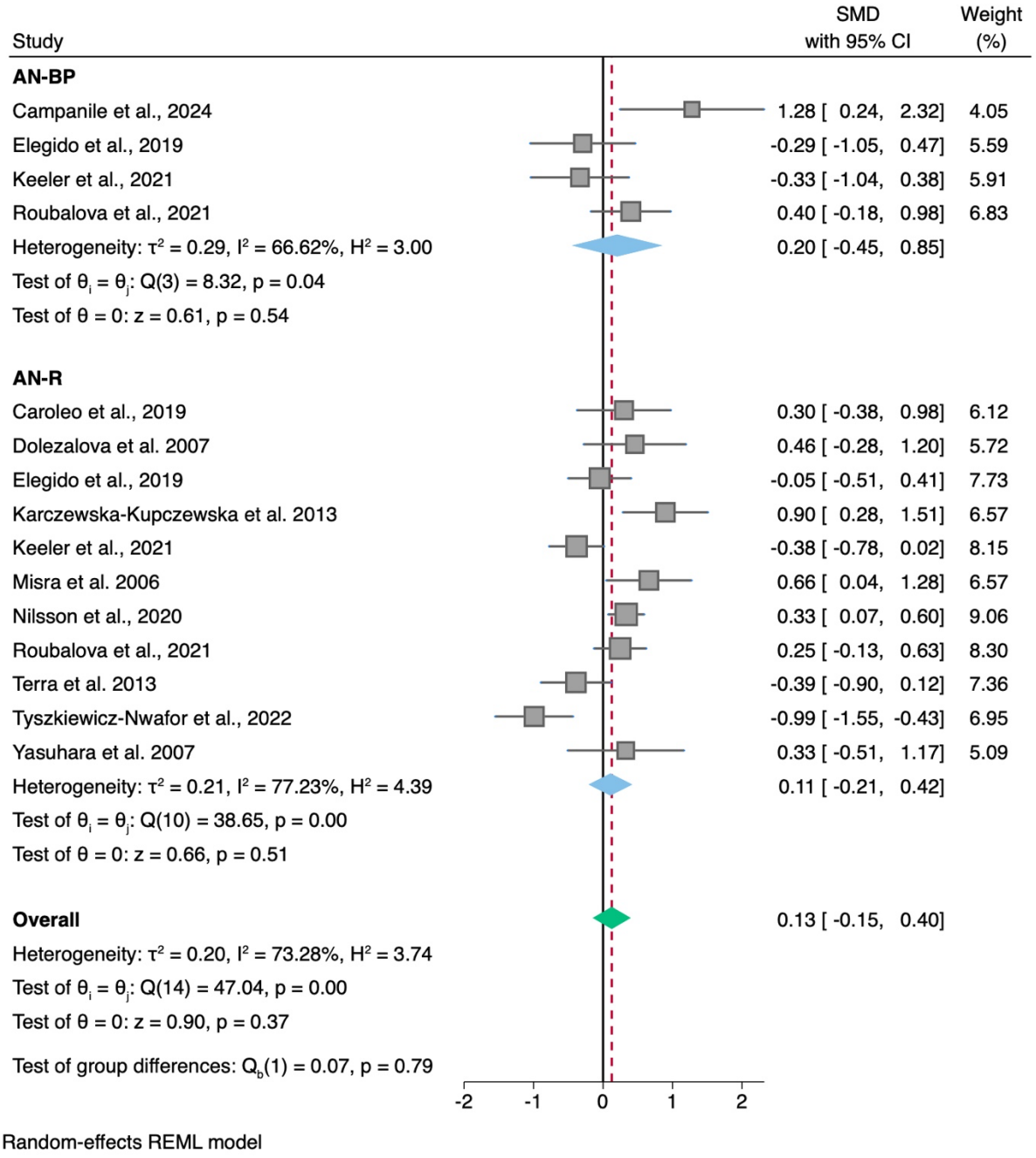

Figure S26. Forest plot of standardized mean difference in interleukin-6 (IL-6) between anorexia nervosa (AN) participants and healthy controls stratified by AN subtype. Zero is the line of no effect, and points to the right of zero indicate an elevation of the cytokine in AN compared to HCs. Abbreviations: AN-BP = AN binge-purge subtype; AN-R = AN restricting subtype; CI = confidence intervals; SMD = standardized mean difference.

5.2. Cross-sectional meta-analyses of differences in concentrations of additional cytokines between anorexia nervosa and healthy controls

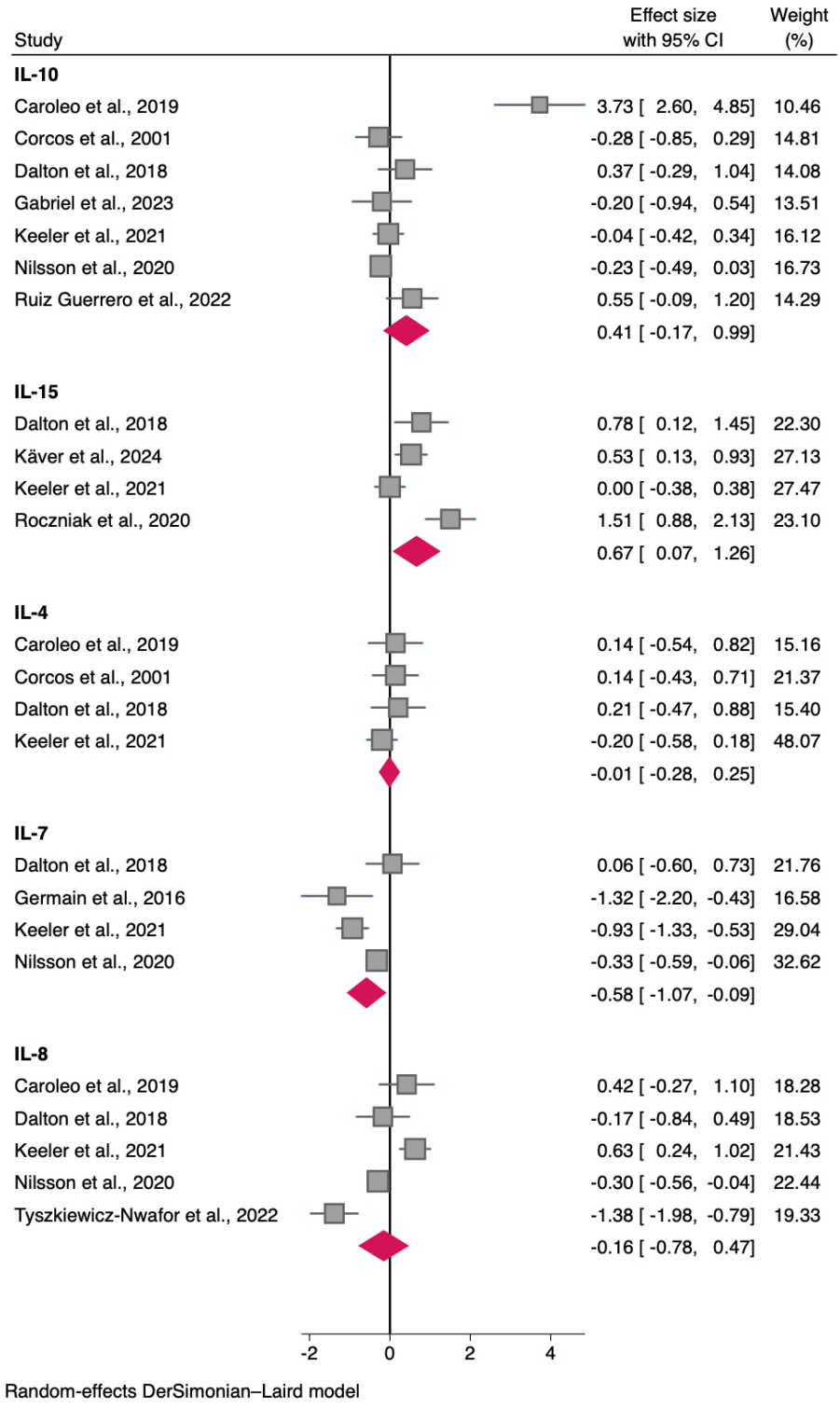

Figure S27. Forest plot of standardized mean difference in interleukin (IL)-4, -7, -8, -10 and -15 between AN participants and HCs. Zero is the line of no effect, and points to the right of zero indicate an elevation of the cytokine in AN compared to HCs. Abbreviations: CI = confidence intervals; IL = interleukin; SMD = standardized mean difference.

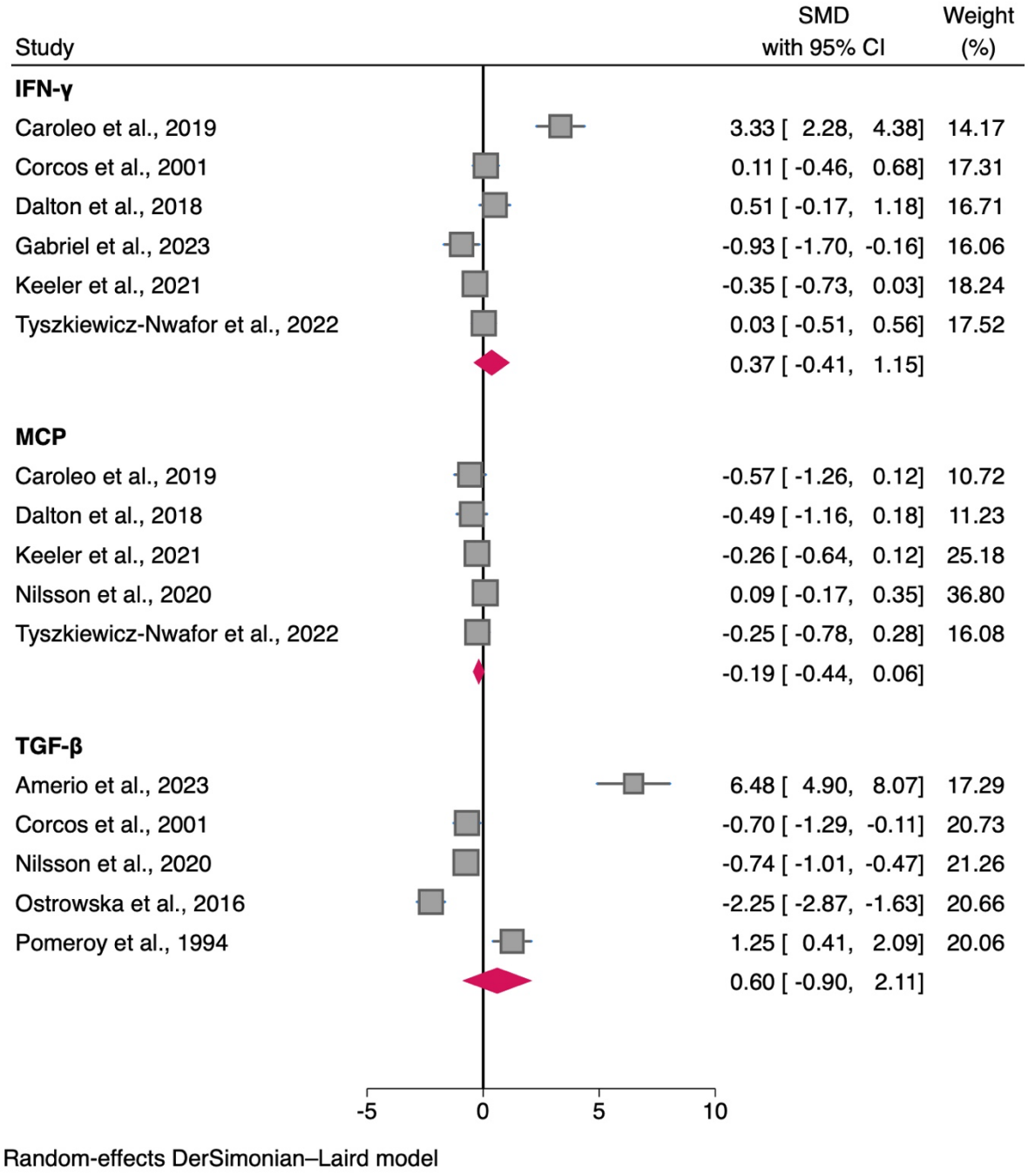

Figure S28. Forest plot of standardized mean difference in interferon-gamma (IFN-γ), monocyte chemoattractant protein-1 (MCP) and transforming growth factor-beta (TGF-β) between AN participants and HCs. Zero is the line of no effect, and points to the right of zero indicate an elevation of the cytokine in AN compared to HCs. Abbreviations: CI = confidence intervals; IFN-g = interferon-gamma; MCP = monocyte chemoattractant protein-1; SMD = standardized mean difference; TGF-b = transforming growth factor-beta.

5.3. Meta-analyses of differences in cytokine concentrations in participants with anorexia nervosa between baseline and follow-up

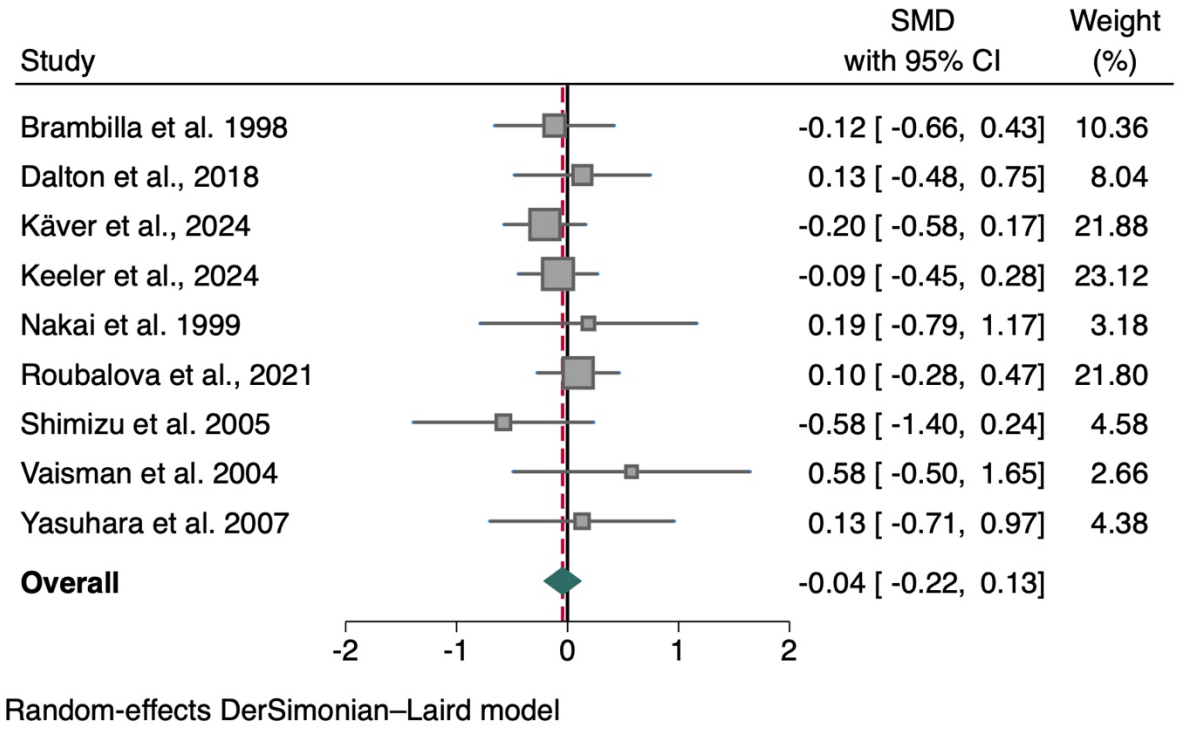

Figure S29. Forest plot of standardized mean difference in tumour necrosis factor-alpha (TNF-α) between AN participants at baseline and follow-up. Zero is the line of no effect, and points to the right of zero indicate an elevation of the cytokine at baseline compared to follow-up. Abbreviations: CI = confidence intervals; SMD = standardized mean difference.

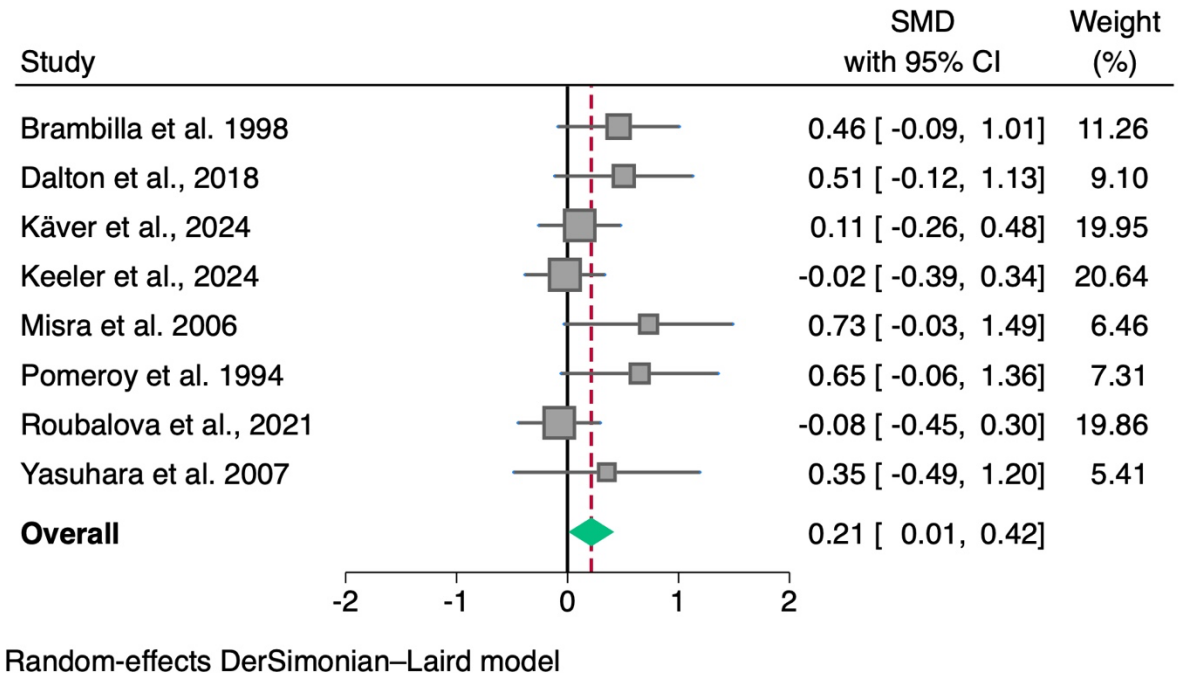

Figure S30. Forest plot of standardized mean difference in interleukin-6 (IL-6) between AN participants at baseline and follow-up. Zero is the line of no effect, and points to the right of zero indicate an elevation of the cytokine at baseline compared to follow-up. Abbreviations: CI = confidence intervals; SMD = standardized mean difference.

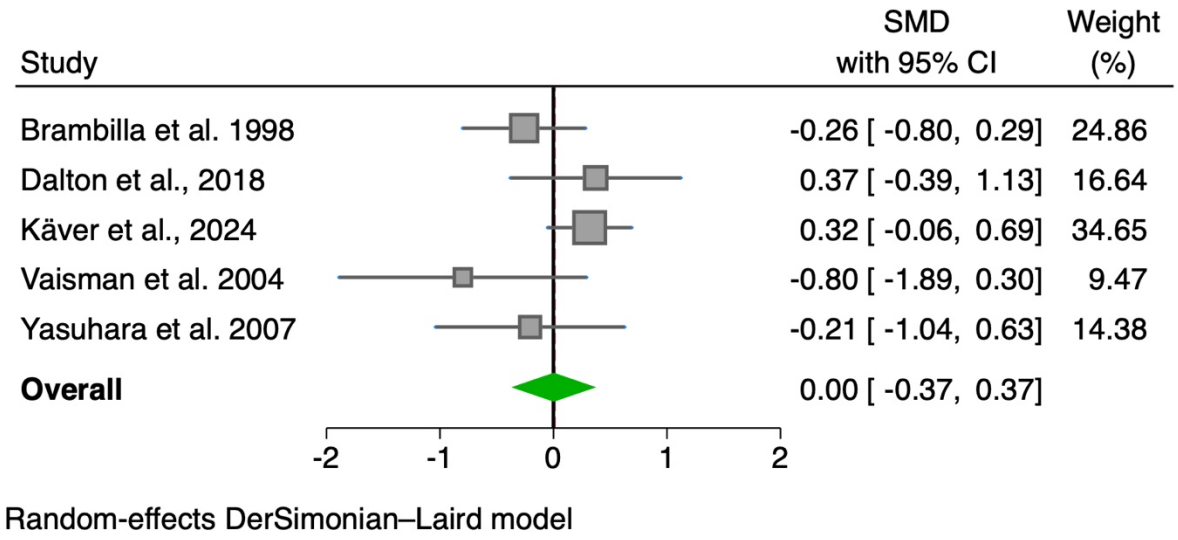

Figure S31. Forest plot of standardized mean difference in interleukin-1β (IL-1β) between AN participants at baseline and follow-up. Zero is the line of no effect, and points to the right of zero indicate an elevation of the cytokine at baseline compared to follow-up. Abbreviations: CI = confidence intervals; SMD = standardized mean difference.

6. MetaForest methodology

First, the approach ranks moderators in terms of their importance in predicting the effect size; recursive preselection plots are seen in Figures S32-33. Using 100-fold replicated feature selection, moderators that reduced predictive performance (i.e., those with positive variable importance in fewer than 50% of replications) were excluded. The remaining moderators were then used to optimize the model (model tuning), consisting of 10,000 regression trees with random-effects weights, again with 100 replications to ensure robustness. The model's predictive performance was evaluated using the "out-of-bag" measure (R^2_{oob}), which estimates the variance explained by the moderators in a new dataset. A negative R^2_{oob} value indicates that the model's predictions do not improve over simply using the average effect size as a predictor, suggesting that the moderators do not have a significant association with the effect size. Positive values indicate that the moderators explain variance in new data, and the most important moderators were then entered into meta-regressions.

6.1. MetaForest moderator analysis for meta-analysis of TNF-α concentrations between anorexia nervosa and healthy controls

To further investigate sources of heterogeneity, a random-effects MetaForest analysis was conducted. Irrelevant moderators were identified using 100-fold replicated feature selection, with only those showing positive variable importance in more than 50% of replications carried forward. Five moderators were retained for model tuning (Figure S32). The main analysis consisted of 10,000 regression trees with random-effect weights, one candidate variable per split and a minimum of 10 cases per terminal node. The final model showed negative estimates of explained variance in new data (R_{oob}^2 = -0.03; R_{cv}^2 = 0.61), indicating that the average effect size predicted new data better than the model, suggesting no significant associations between the moderators and the SMD. This remained unchanged when study outliers were removed, with eight moderators carried forward at the preselection stage (R_{oob}^2 = -0.01; R_{cv}^2 = 0.78).

6.2. MetaForest moderator analysis for meta-analysis of IL-6 concentrations between anorexia nervosa and healthy controls

Using MetaForest to further explore sources of heterogeneity, six studies were retained for model tuning. The final model showed negative estimates of explained variance in new data (R_{oob}^2 = -0.06; R_{cv}^2 = 0.69), indicating no significant associations between the candidate moderators and the SMD for TNF-α. However, after removing study outliers, 11 moderators were carried forward at the preselection stage (replicated variable importance metrics shown in Figure S33). An identical main analysis, using 11 candidate variables per split and a minimum of three cases per terminal node, produced positive estimates of explained variance in new data

(R_{oob}^2 = 0.13; R_{cv}^2 = 0.68). The final model identified the following moderators as the most important in explaining the effect size: study year, NOS score, study region, age group, percentage of sample using medication, mean age, BMI, and illness duration of the AN sample, and mean age and BMI of the HC sample. The relative variable importance of these moderators is shown in Figure S34. The marginal relationship of each moderator with the effect size, averaging over all values of other moderators, is illustrated in Figure S35. These moderators were entered into meta-regressions (Table S2). Only study country was significant, with studies conducted in North America showing a greater difference between AN and HC compared to studies in Europe (B = 0.78; SE = 0.29; *z* = 2.72; *p* = 0.007; 95% CI 0.22, 1.34).

6.3. MetaForest moderator analysis for meta-analysis of IL-1β concentrations between anorexia nervosa and healthy controls

To investigate the remaining sources of heterogeneity, a random effects MetaForest analysis was again utilised, using identical parameters. One moderator was carried forward for model tuning. Again, the final model had negative estimates of explained variance in new data (R_{oob}^2 = -0.14). When removing the three study outliers, no moderators were identified for moderator tuning. Therefore, there was no evidence for associations between the entered moderators and the SMD for IL-1β.

7. MetaForest plots

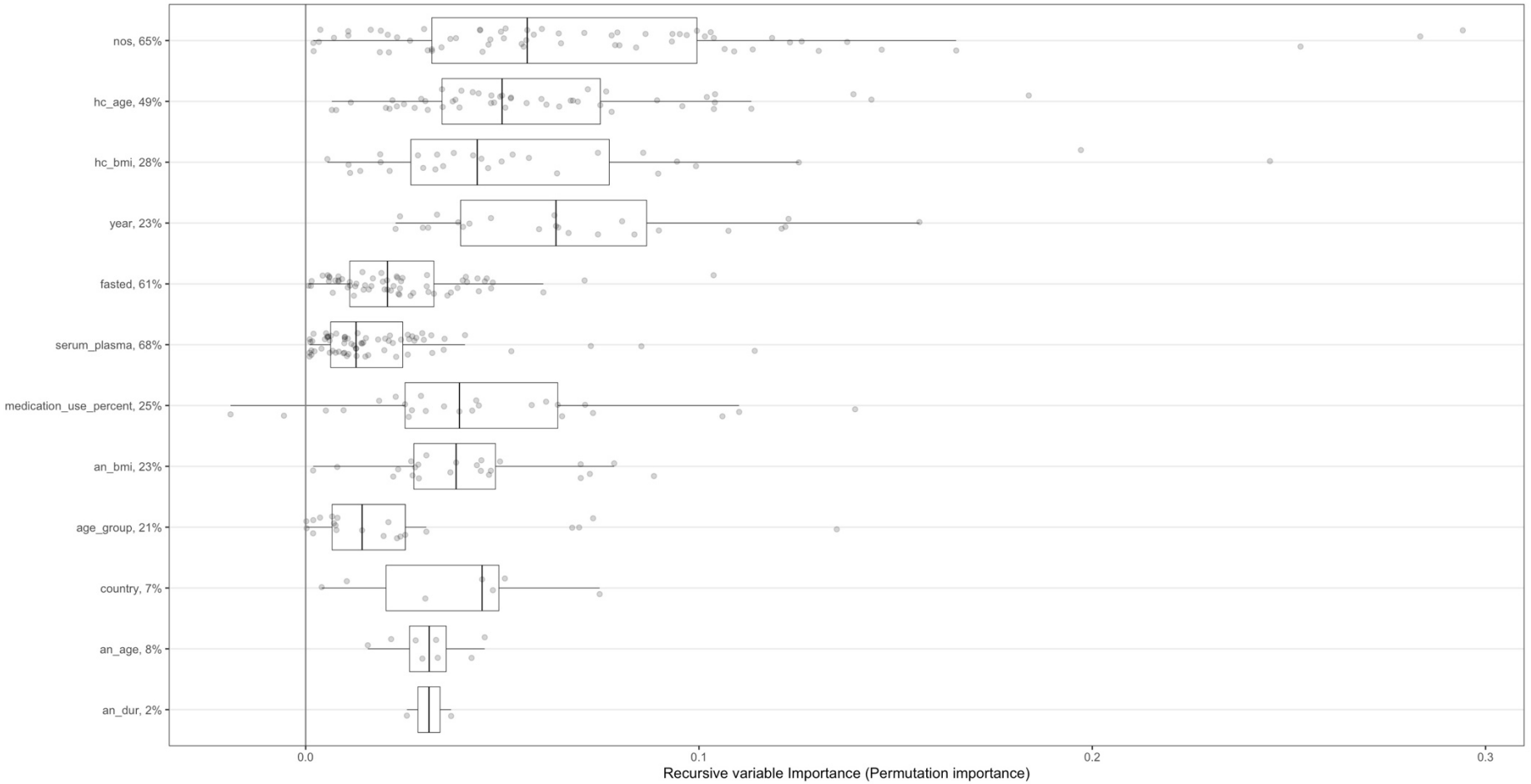

Figure S32. Replicated MetaForest for variable preselection, meta-analysis of tumour necrosis factor-alpha (TNF-α) concentrations between anorexia nervosa and healthy controls.

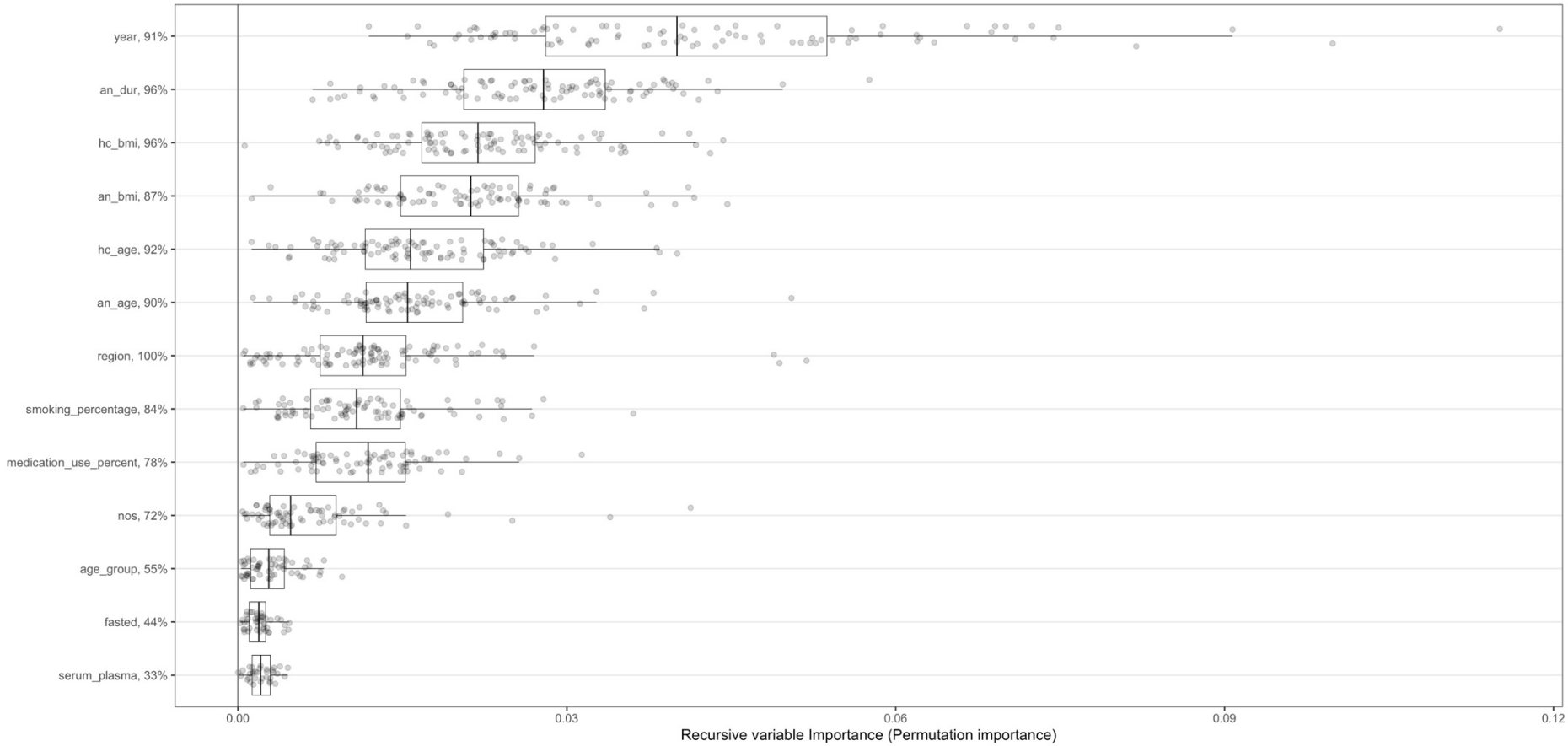

Figure S33. Replicated MetaForest for variable preselection, meta-analysis of interleukin-6 (IL-6) concentrations between anorexia nervosa and healthy controls after removing study outliers.

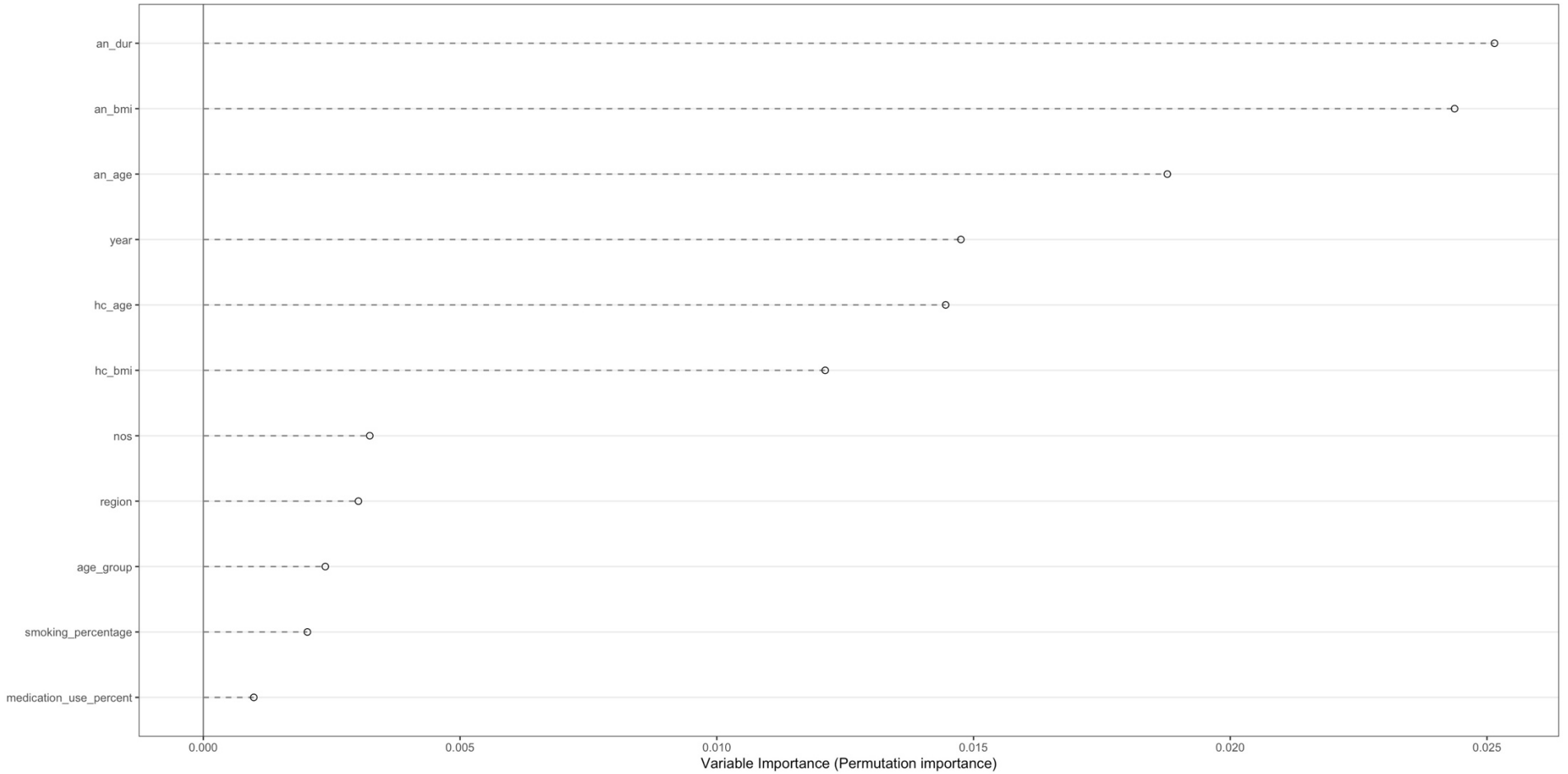

Figure S34. Variable importance for final model, meta-analysis of interleukin-6 (IL-6) concentrations between anorexia nervosa and healthy controls after removing study outliers.

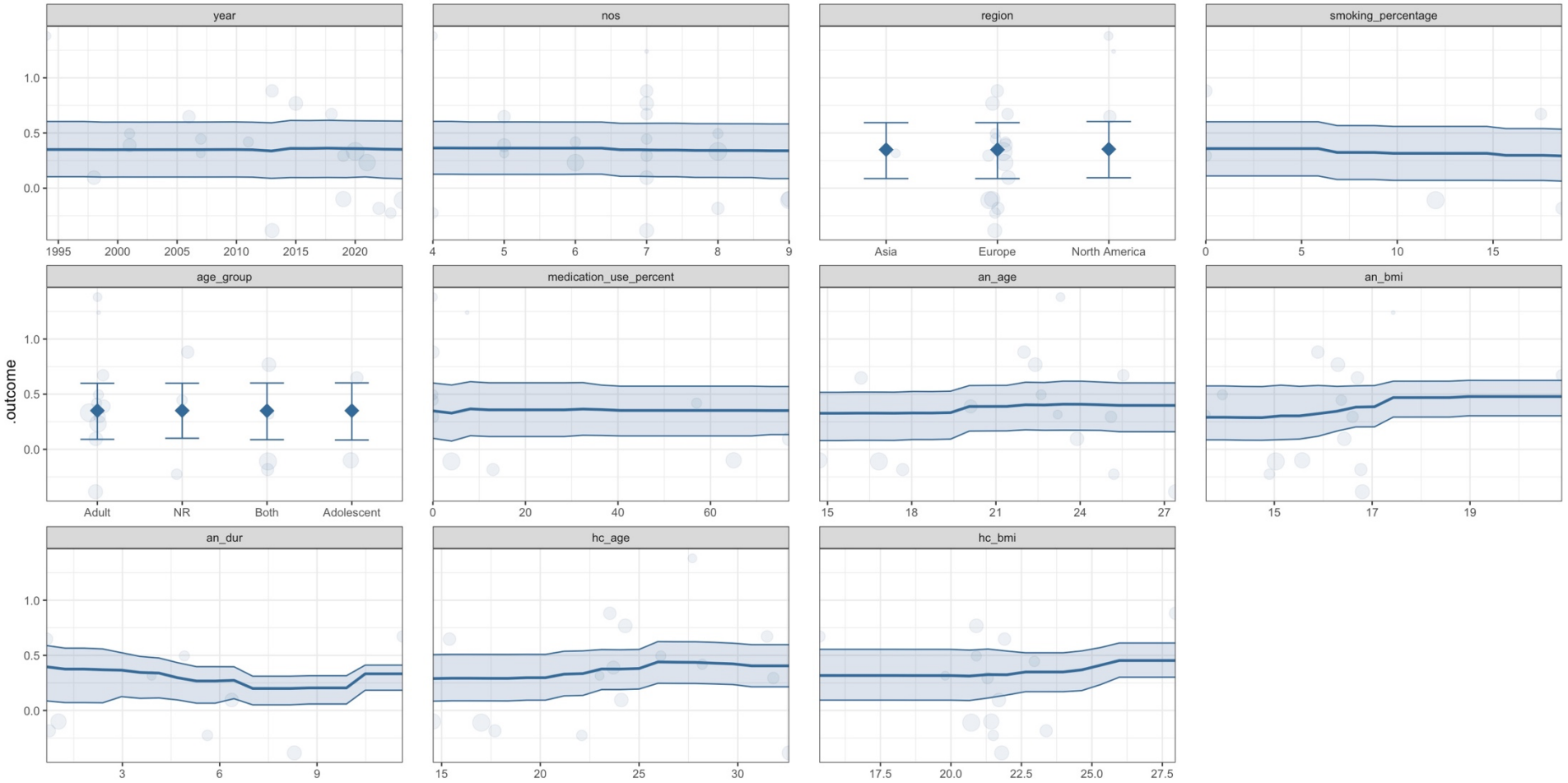

Figure S35. Partial dependence plots, meta-analysis of interleukin-6 (IL-6) concentrations between anorexia nervosa and healthy controls after removing study outliers.

8. Exploratory meta regressions based on results of MetaForest moderator analyses

Table S2. Results of exploratory individual meta-regressions for moderators selected from MetaForest analysis on standardised mean differences in interleukin-6 (IL-6) concentrations between anorexia nervosa and healthy controls, after removing five study outliers (Amerio et al., 2023; Käver et al., 2024; Keeler et al., 2021; Ostrowska et al., 2015; Tyszkiewicz-Nwafor et al., 2022).

Variable	Estimate	SE	z-value	p-value	(95% CI)
Study year	-0.02	0.01	-1.78	0.074	-0.04, 0.002
Study NOS score	-0.11	0.06	-1.77	0.077	-0.23, 0.01
Study region*Asia[a]	0.10	0.50	0.19	0.848	-0.88, 1.08
Study region*North America[a]	0.78	0.29	2.72	0.007**	0.22, 1.34
Percentage of sample who smoke	-0.03	0.03	-0.99	0.324	-0.08, 0.03
Age group*Adults[b]	0.15	0.32	0.47	0.636	-0.48, 0.79
Age group*Both[b]	-0.05	0.36	-0.15	0.884	-0.77, 0.66
Percentage of sample using psychotropic medication	-0.01	0.01	-1.28	0.202	-0.02, 0.003
Mean age (AN)	0.01	0.03	0.39	0.695	-0.05, 0.08
Mean BMI (AN)	0.08	0.08	1.01	0.311	-0.07, 0.23
Mean illness duration (AN)	0.01	0.04	0.23	0.821	-0.07, 0.09
Mean age (HC)	0.02	0.02	0.83	0.406	-0.02, 0.06
Mean BMI (HC)	0.02	0.05	0.24	0.809	-0.08, 0.11

[a]Reference category was Europe. [b]Reference category was adolescents. Abbreviations: AN = anorexia nervosa; BMI = body mass index; CI = confidence intervals; HC = healthy controls; NOS = Newcastle-Ottawa Scale; SE = standard error.

9. Results of additional studies

Table S3. Results of additional studies not included in meta-analysis

Cytokine	AN vs HC	RecAN vs HC	BN vs HC	BED vs HC	AN longitudinal
Eotaxin	↑ (Keeler et al., 2021; Nilsson et al., 2020) ↔ (Dalton et al., 2018)	↔ (Keeler et al., 2021; Nilsson et al., 2020)			↔ (Dalton et al., 2020)
Eotaxin-3	↔ (Dalton et al., 2018) ↑ (Keeler et al., 2021)	↓ (Keeler et al., 2021)			↔ (Dalton et al., 2020)
GM-CSF	↔ (Dalton et al., 2018; Keeler et al., 2021)	↔ (Keeler et al., 2021)			↔ (Dalton et al., 2020; Vaisman et al., 2004)
IFN-γ	MA	↔ (Keeler et al., 2021)			↔ (Dalton et al., 2020) ↑ (after 10 week diet including yogurt; Solis et al., 2002) ↓ (after 10 week diet including milk; Solis et al., 2002)
IL-1	↔ (Corcos et al., 2001; Vaisman et al., 1996)				
IL-1ß	MA	MA	↑ (Tabasi et al., 2020) ↔ (Brambilla et al., 1998; Monteleone et al., 1999)	↔ (Caroleo et al., 2019	MA
IL-1α	↑ (Caroleo et al., 2019) ↔ (Dalton et al., 2018)	ND (Nilsson et al., 2020)		↓ (Caroleo et al., 2019)	↔ (Dalton et al., 2020)
IL-2	↔ (Caroleo et al., 2019; Dalton et al., 2018; Keeler et al., 2021; Allende et al., 1998) ↓ (Corcos et al., 2001)	↔ (Keeler et al., 2021)		↔ (Caroleo et al., 2019)	↔ (Dalton et al., 2020)
IL-3	ND (Vaisman et al., 1996)				↔ (Vaisman et al., 2004)
IL-4	MA	↔ (Keeler et al., 2021)	↔ (Tabasi et al., 2020)	↔ (Caroleo et al., 2019)	↔ (Dalton et al., 2020)

IL-5	↔ (Dalton et al., 2018; Allende et al., 1998)	ND (Nilsson et al., 2020)			↔ (Dalton et al., 2020)
IL-6	MA	MA	MA	↔ (Caldas et al., 2022)	
IL-7	MA	↔ (Keeler et al., 2021) ↓ (Germain et al., 2016)	↔ (Germain et al., 2016)		↑ Week 24 (Dalton et al., 2020)
IL-8	MA	↔ (Keeler et al., 2021; Nilsson et al., 2020)		↔ (Caroleo et al., 2019)	↔ (Dalton et al., 2020)
IL-9	↓ (Gabriel et al., 2023)				
IL-10	MA	↔ (Keeler et al., 2021; Nilsson et al., 2020)	↓ (Tabasi et al., 2020)	↓ (Caroleo et al., 2019)	↔ (Dalton et al., 2020)
IL-12/IL-23p40	↔ (Dalton et al., 2018) ↓ (Keeler et al., 2021)	↔ (Keeler et al., 2021)			↔ (Dalton et al., 2020)
IL-12B	↓ (Nilsson et al., 2020; Breithaupt et al., 2024)	↔ (Nilsson et al., 2020)			
IL-12p70	↔ (Dalton et al., 2018; Keeler et al., 2021)	↑ (Keeler et al., 2021)			↔ (Dalton et al., 2020)
IL-13	↔ (Dalton et al., 2018; Keeler et al., 2021)	↔ (Keeler et al., 2021)	↓ (Tabasi et al., 2020)		↔ (Dalton et al., 2020)
IL-15	MA				↔ (Dalton et al., 2020)
IL-16	↔ (Dalton et al., 2018) ↓ (Keeler et al., 2021)				↔ (Dalton et al., 2020)
IL-17A	↔ (Breithaupt et al., 2024; Dalton et al., 2018; Roubalova et al., 2021) ↓ (Keeler et al., 2021; Gabriel et al., 2023)	↓ (Keeler et al., 2021)			↔ (Dalton et al., 2020; Roubalova et al., 2021)
IL-17C	↔ (Nilsson et al., 2020)	↔ (Nilsson et al., 2020)			
IL-17F	↓ (Gabriel et al., 2023)				
IL-18	↔ (Breithaupt et al., 2024) ↓ (Nilsson et al., 2020)	↔ (Nilsson et al., 2020)			
IL-20	ND (Nilsson et al., 2020)	ND (Nilsson et al., 2020)			
IL-21	↑ (Amerio et al., 2023)				
IL-22	↔ (Gabriel et al., 2023; Keeler et al., 2021)	↔ (Keeler et al., 2021)			

IL-23	↑ (Gabriel et al., 2023)				
IL-24	ND (Nilsson et al., 2020)	ND (Nilsson et al., 2020)			
IL-25	↓ (Gabriel et al., 2023)				
IL-27	↔ (Keeler et al., 2021)	↔ (Keeler et al., 2021)			
IL-31	↔ (Keeler et al., 2021) ↓ (Gabriel et al., 2023)	↔ (Keeler et al., 2021)			
IL-33	↔ (Gabriel et al., 2023	ND (Nilsson et al., 2020)			
IP-10	↔ (Dalton et al., 2018; Keeler et al., 2021)	↔ (Keeler et al., 2021)			↔ (Dalton et al., 2020)
MCP-1	MA	↔ (Nilsson et al., 2020) ↓ (Keeler et al., 2021)			↔ (Dalton et al., 2020)
MCP-2	↔ (Breithaupt et al., 2024; Nilsson et al., 2020)	↔ (Nilsson et al., 2020)			
MCP-3	↓ (Nilsson et al., 2020)	↔ (Nilsson et al., 2020)			
MCP-4	↔ (Breithaupt et al., 2024; Dalton et al., 2018; Nilsson et al., 2020)	↔ (Nilsson et al., 2020)			↔ (Dalton et al., 2020)
MDC	↔ (Keeler et al., 2021)	↓ (Keeler et al., 2021)			
MIP-1ß	↔ (Breithaupt et al., 2024; Dalton et al. 2018; Nilsson et al., 2020) ↓ (Keeler et al., 2021)	↔ (Nilsson et al., 2020) ↓ (Keeler et al., 2021)			↔ (Dalton et al., 2020)
MIP-1α	↔ (Dalton et al. 2018; Keeler et al., 2021; Nilsson et al., 2020)	↔ (Nilsson et al., 2020) ↓ (Keeler et al., 2021)			↔ (Dalton et al., 2020)
MIP-3α	↔ (Keeler et al., 2021)	↔ (Keeler et al., 2021)			
TARC	↔ (Dalton et al., 2018; Keeler et al., 2021)	↔ (Keeler et al., 2021)			↔ (Dalton et al., 2020)
TGF-ß	MA	↔ (Nilsson et al., 2020)	↔ (Tabasi et al., 2020)		
TNF-ß	↓ (Dalton et al., 2018; Nilsson et al., 2020) ↔ (Breithaupt et al., 2018; Keeler et al., 2021)	↔ (Keeler et al., 2021; Nilsson et al., 2020)			↔ (Dalton et al., 2020)
TNF-α	MA	↓ (Keeler et al., 2021)	MA	↑ (Caldas et al., 2022)	MA

Abbreviations: AN = anorexia nervosa; BED = binge eating disorder; BN = bulimia nervosa; GM-CSF = granulocyte-macrophage colony-stimulating factor; HC = healthy controls; IFN-γ = interferon-gamma; IL = interleukin; IP-10 = interferon gamma-induced protein-10; MA = meta-analysed in main manuscript; MCP = monocyte chemotactic protein; MDC = macrophage-derived chemokine; MIP = macrophage inflammatory protein; ND = not detectable; recAN = recovered AN; TARC = thymus and activation-regulation chemokine; TGF = transforming growth factor; TNF = tumour necrosis factor.

9.1. Narrative summary of additional studies not included in meta-analysis

Anorexia nervosa versus healthy controls. In people with AN compared to controls, concentrations of IL-21 and IL-23 were found to be increased, and concentrations of Eotaxin, Eotaxin-3 and IL-1α were increased in some studies but were similar to controls in other studies. Cytokines found to be decreased in AN included IL-9, IL-12B, IL-17F, IL-23, and monocyte chemotactic protein (MCP)-3. There were mixed evidence for decreases in AN in some studies and no difference in comparison to controls in other studies, in concentrations of IL-2, IL-12/IL-23p40, IL-16, IL-17A, IL-18, IL-31, MIP-1ß and TNF-ß. Additionally, the following cytokines were found to not differ between AN and controls: granulocyte-macrophage colony-stimulating factor (GM-CSF), IL-1, IL-5, IL-12p70, IL-13, IL-17C, IL-22, IL-27, IL-33, IP-10, MCP-2, MCP-4, macrophage-derived chemokine (MDC), MIP-1α, MIP-3α, and thymus and activation-regulation chemokine (TARC).

Anorexia nervosa baseline to follow-up. Longitudinally, concentrations of 23 cytokines were found not to change, although notably these findings were largely from a study whereby full weight gain did not occur longitudinally (Dalton et al., 2020). These cytokines included: Eotaxin, Eotaxin-3, GM-CSF, IL-1α, IL-2, IL-3, IL-4, IL-5, IL-8, IL-10, IL-12/IL-23p40, IL-12p70; IL-13; IL-15; IL-16; IL-17A, IP-10, MCP-1, MCP-4, MIP-1ß, MIP-1α, TARC and TNF-ß. This same study found a longitudinal increase in IL-7 after a 24-week follow-up. For interferon (IFN)-γ, there were mixed results whereby one study found no longitudinal change, and another study found an increase after 10 weeks of nutritional reinstatement including yogurt intake, but a decrease after 10 weeks of nutritional reinstatement including milk intake.

Individuals recovered from anorexia nervosa versus healthy controls. People recovered from AN (recAN) showed increased concentrations of IL-12p70 and decreased concentrations of Eotaxin-3, IL-17A, MDC and TNF-α in comparison with healthy controls. Moreover, for IL-7, MCP-1, MIP-1α and MIP-1ß, some studies found decreased concentrations in recAN and some found no difference between recAN and controls. Concentrations of 23 other cytokines were similar between recAN and controls, which included: Eotaxin, GM-CSF, IFN-γ, IL-2, IL-4, IL-8, IL-10, IL-12/IL-23p40, IL-12B, IL-13, IL-17C, IL-18, IL-22, IL-27, IL-31, IP-10, MCP-2, MCP-3, MCP-4, MIP-3α, TARC, transforming growth factor (TGF)-ß, and TNF-ß.

Bulimia nervosa versus healthy controls. Concentrations of IL-10 and IL-13 were found to be decreased in people with BN compared to controls. Concentrations of IL-4, IL-7, TGF-ß and IL-1ß were found to be similar between BN and controls, although one study found an increase in IL-1ß in people with BN.

Binge eating disorder versus healthy controls. In BED, it was found that concentrations of IL-10 were decreased in comparison to controls, and concentrations of TNF-α were increased. Concentrations of IL-1α, IL-1ß, IL-2, IL-4, IL-6 and IL-8 were comparable between BED and controls. However, all findings were based on one study only.

Otherwise-specified feeding and eating disorder versus healthy controls. One study measured IL-6 concentrations in women with OSFED, which was numerically higher than controls (SMD = 0.52), although inferential statistics for this comparison were not presented (Campanile et al., 2024).