Article

Data extraction errors in meta-analyses impact the association between depression and peripheral inflammatory biomarkers: an umbrella review

San Lee[1,2,3], Kyung Mee Park[1,2], Sung Jong Park[1], Won Jae Kim[1], Keum Hwa Lee[4], Jinhee Lee[5], Andreas Kronbichler[6], Lee Smith[7], Marco Solmi[8,9,10], Brendon Stubbs[11,12], Ai Koyanagi[13,14], Louis Jacob[13,15], Andrew Stickley[16], Trevor Thompson[17], Elena Dragioti[18], Hans Oh[19], Andre R. Brunoni[20], Andre F. Carvalho[21], Joaquim Radua[10,22,23], Suk Kyoon An[1,2], Kee Namkoong[1,2], Eun Lee[1,2,*], Jae Il Shin[4,*], and Paolo Fusar-Poli[10,24,25]

[1] Department of Psychiatry, Yonsei University College of Medicine, Seoul, Republic of Korea

[2] Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

[3] Department of Psychiatry, Yongin Severance Hospital, Yonsei University College of Medicine, Yongin, Republic of Korea

[4] Department of Pediatrics, Yonsei University College of Medicine, Seoul, Republic of Korea

[5] Department of Psychiatry, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea

[6] Department of Internal Medicine IV (Nephrology and Hypertension), Medical University Innsbruck, 6020 Innsbruck, Austria

[7] The Cambridge Centre for Sport and Exercise Sciences, Anglia Ruskin University, Cambridge CB1 1PT, UK

[8] Department of Neuroscience, University of Padua, Padua, Italy.

[9] Padova Neuroscience Center (PNC), University of Padua, Padua, Italy

[10] Early Psychosis: Interventions and Clinical-Detection (EPIC) Lab, Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[11] Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, London SE5 8AF, UK

[12] Physiotherapy Department, South London and Maudsley NHS Foundation Trust, Denmark Hill, London SE5 8AZ, UK

[13] Parc Sanitari Sant Joan de Déu/CIBERSAM, Universitat de Barcelona, Fundació Sant Joan de Déu, Sant Boi de Llobregat, Barcelona, Spain

[14] ICREA, Pg. Lluis Companys 23, Barcelona, Spain

[15] Faculty of Medicine, University of Versailles Saint-Quentin-en-Yvelines, Montigny-le-Bretonneux, France

[16] Stockholm Center for Health and Social Change (SCOHOST), Södertörn University, Huddinge 141 89, Sweden

[17] Department of Psychology, University of Greenwich, London SE109LS, UK

[18] Pain and Rehabilitation Centre, and Department of Medical and Health Sciences, Linköping University, SE-581 85 Linköping, Sweden

[19] School of Social Work, University of Southern California, CA, USA

[20] Department of Psychiatry and Psychotherapy, University Hospital, LMU Munich, Munich, Germany; Service of Interdisciplinary Neuromodulation, Department of Psychiatry, Laboratory of Neurosciences (LIM-27) and National Institute of Biomarkers in Neuropsychiatry (INBioN), Institute of Psychiatry, University of Sao Paulo, Sao Paulo, Brazil; Hospital Universitario, Departamento de Clínica Médica, Faculdade de Medicina da USP, São Paulo, Brazil

[21] Centre for Addiction and Mental Health (CAMH), Toronto, ON, Canada; Department of Psychiatry, University of Toronto, Toronto, ON, Canada

[22] Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden

[23] Imaging of Mood- and Anxiety-Related Disorders (IMARD) Group, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain

[24] OASIS service, South London and Maudsley NHS Foundation Trust, London, UK

[25] Department of Brain and Behavioural Sciences, University of Pavia, Pavia, Italy

*** Corresponding Authors**

Eun Lee, MD, PhD, Professor

Department of Psychiatry, Yonsei University College of Medicine, 50, Yonsei-ro, Seodaemun-gu, Seoul, 03722, Republic of Korea

Tel.: +82-2-2228-1620, Fax: +82-2-313-0891, Email: leeeun@yuhs.ac

Jae Il Shin, MD, PhD, Professor
Department of Pediatrics, Yonsei University College of Medicine, 50, Yonsei-ro, Seodaemun-gu, Seoul, 03722, Republic of Korea
Tel.: +82-2-2228-2050, Fax: +82-2-393-9118, Email: shinji@yuhs.ac

Summary

Background: Accumulating evidence indicates that alterations in inflammatory biomarkers are crucially involved in depression. However, previous meta-analyses disagree on the consistency of these associations. Errors in data extraction may explain these discrepancies.

Methods: This umbrella review (PROSPERO: CRD42019133888) included meta-analyses exploring the association between depression and four peripheral inflammatory biomarkers: tumor necrosis factor-alpha (TNF-α), interleukin 1-beta (IL-1β), IL-6, and C-reactive protein (CRP). PubMed/MEDLINE, Embase, PsycINFO, and the Cochrane Library were searched since database inception to Jan 14, 2020, to assess data extraction errors and their impact on the results. The meta-analytical associations were re-estimated after correcting the errors.

Findings: In 14 (93·3%) of the 15 meta-analyses included, there were some errors. Across 521 primary studies, 118 (22·6%) showed some errors: incorrect sample sizes (16·9%), incorrect use of standard deviation (29·7%), incorrect participant inclusion (5·9%), calculation errors (28·0%) and analysis with insufficient data (19·5%). In 305 overlapping primary studies, 61 (20·0%) of them showed some errors. After correcting these errors, 11 (29·7%) out of 37 pooled effect sizes changed the magnitude level of the effect size of more than 0·1. The updated meta-analyses showed that elevated levels of peripheral TNF-α, IL-6, CRP, but not IL-1β, were associated with depression.

Interpretation: Data extraction in meta-analyses can lead to significant errors that impact core findings. Efforts to reduce the errors are important in such studies for the association between depression and peripheral inflammatory biomarkers where high heterogeneity and conflicting results have been reported continuously.

Funding: The National Research Foundation of Korea (2017R1A2B3008214).

Introduction

A growing body of evidence indicates that alterations in immune-inflammatory pathways play important roles in the pathophysiology of depression.[1,2] Compared with healthy controls, patients with depression show elevated blood levels of inflammatory cytokines such as tumor necrosis factor-alpha (TNF-α), interleukin 1-beta (IL-1β), and IL-6.[3,4] In addition, C-reactive protein (CRP), an acute-phase reactant, is elevated in depression.[5]

Given the number of primary studies reporting conflicting results on the associations between inflammatory biomarkers and depression, meta-analyses are typically employed as state-of-the-art empirical summary, integrating data across multiple primary studies and providing a more reliable answer to a research question than a single study.[6] Several meta-analyses have elucidated the associations between depression and inflammatory biomarkers. However, these meta-analyses report inconsistent findings.[3,4,7-13] Although elevated peripheral levels of TNF-α was associated with depression in some meta-analyses,[4,8-10,13] others reported no association between depression and TNF- α.[7,11] Likewise, IL-1β was associated with depression in several meta-analyses,[3,7] but not in others.[4,8-11,13]

Errors in data extraction can be one explanation for why different meta-analyses reach different conclusions for the same research question. Investigators who have attempted to replicate published meta-analyses found that 59%–100% contain errors.[14-16] Such errors in data extraction can lead to overestimate or nullify the significance of the results. For example, when performing a meta-analysis, it is sometimes necessary to standardize measurements on a uniform scale, such as standardized mean difference (SMD), before pooling across primary studies. During this process, sample size may be incorrectly exported[17] and standard errors (SEs) may be mistaken for standard deviations (SDs), which can substantially inflate point estimates and heterogeneity.[15] Inaccuracy in calculation[18] and data analysis with incomplete

information[19] is also reported during data extraction.

To address this, we selected four peripheral inflammatory biomarkers, which have been extensively investigated for the association with depression: TNF-α, IL-1β, IL-6, and CRP. Then, we examined errors in meta-analyses of the association between depression and the four peripheral inflammatory biomarkers. We employed an umbrella review of meta-analyses to evaluate the presence, frequency, and nature of errors in data extraction and their impact on the results. Furthermore, we corrected the errors and then re-estimated the meta-analytical association between depression and the peripheral inflammatory biomarkers. Finally, we collected all primary studies included in the meta-analyses and calculated the updated total pooled effect sizes (ESs) of the association between depression and the inflammatory biomarkers. With the updated total pooled ESs, we aimed to evaluate the association of immune-inflammatory pathway with depression.

Methods

Search strategy and selection criteria

Four investigators (SL, KMP, SJP, and WJK) searched PubMed/MEDLINE, Embase, PsycINFO, and the Cochrane Library for articles published between database inception and January 14, 2020, using the search terms (CRP OR IL-1beta OR IL-6 OR TNF-alpha) AND depress* AND meta using the [All Fields] search tag for all terms. The searching process was done until February 11, 2019, and then repeated until January 14, 2020, to update newly published meta-analyses. The full names and abbreviations of all four peripheral inflammatory biomarkers were employed in the search strategy. We chose eligible articles by consecutively screening their titles and abstracts followed by their full texts (figure 1). Disagreements were resolved via discussion among the authors SL, KHL, EL, and JIS.

We included meta-analyses of observational studies examining the association between unipolar or bipolar depression[9,20-22] and levels of TNF-α, IL-1β, IL-6, or CRP in circulating blood (plasma/serum). Some meta-analyses included primary studies for any depressive disorder and others included in the study only for studies on major depressive disorders. Our definition of depression followed that of each original meta-analysis. The international prospective register of systematic reviews (PROSPERO) registration status was evaluated.

We screened articles without language restriction to avoid language restriction bias. We only included meta-analyses that reported ESs for individual primary studies or the data necessary for their calculation. We use the term 'overlapping meta-analyses' to indicate meta-analyses of the same association between depression and an inflammatory biomarker, and we use the term 'overlapping studies' to indicate primary studies that were included in more than one meta-analysis.

Data extraction

From each meta-analysis, four investigators (SL, KMP, SJP, and WJK) extracted the first author, publication date, literature search date, inflammatory biomarker of interest, model of analysis (i.e., fixed effect or random effects), sample sizes, maximally adjusted individual study estimates and corresponding 95% confidence intervals (CIs), and ES metrics presented for results (e.g., SMD (Cohen's *d)*, Hedges' *g*, mean difference, or odds ratio). From the primary studies included in the meta-analyses, we extracted sample size and mean ± SD, mean ± SE or median and interquartile range (IQR) for each inflammatory biomarker. If data were presented in terms of median and IQR, the calculation method for their conversion to mean ± SD was investigated. If a study compared two or more subgroups of depression to the same control, we combine subgroups to create a single pair-wise comparison. When an outcome was measured

in a single population study, the correlation coefficient was used for calculating SMD. In some primary studies reporting stimulated levels of inflammatory biomarkers from *in vitro* assays were found in a meta-analysis.[13] In the cases, extracted data were used in evaluation for the detection of errors and recalculation for ES of the meta-analysis, but not in the calculation of total pooled results from all primary studies. If data were presented only in graphs, they were extracted with GetData Graph Digitizer (version 2.26).[23]

Data analysis

We recalculated ESs and 95% CIs of primary studies included in the meta-analyses and reanalyzed each meta-analysis accordingly. Pooled ESs, 95% CIs, and p-values were recalculated using Comprehensive Meta-Analysis (version 3.3.070, Biostat, Englewood, NJ, USA). The level of statistical significance was set at $p < 0{\cdot}05$.

To evaluate discrepancies between initial results and those re-estimated, we applied 0·1 as a cut-off point according to a previous review, which also evaluated data extraction errors in meta-analyses.[15] In comparison, we followed the ES metrics of the original meta-analysis (i.e., SMD, Hedges' *g*, mean difference, or odds ratio) in recalculation. If our recalculated results for an ES or its CI differed from those of a primary study reported in the meta-analysis by 0·1 or more, this was regarded as an error.

If errors were identified, they were classified as 'incorrect sample sizes', 'incorrectly used SD', 'incorrect participant inclusion', 'calculation error', or 'analysis with insufficient data' as indicated below. We defined a case in which sample sizes in a primary study were wrongly extracted as 'incorrect sample sizes'. If an extracted SD from a primary study was incorrect, the case was regarded as 'incorrectly used SD'. When a primary study that does not meet the inclusion criteria of each meta-analysis was included, the case was defined as 'incorrect

participant inclusion'. 'Calculation error' indicates a case in which reported effect size is inaccurate despite no errors in reported primary study data for calculating effect size. If sufficient information to calculate SMD was not provided in a primary study, the case was classified as 'analysis with insufficient data'.

We conducted data analysis serially. At first, we evaluated the presence and type of errors in all primary studies in each meta-analysis. After that, we repeated to evaluate errors in only 'overlapping studies' that were included in more than one meta-analysis. If data were extracted directly from previous meta-analyses, not from primary studies, and an error in the previous meta-analyses existed, there is a chance of error duplication from previous meta-analyses. Then, we recalculated pooled ESs of the meta-analyses after correcting errors. If there was a case in which an initial pooled ES is different more than 0·1 from our recalculated value, this was presented as a 'change in results'. Lastly, we gathered all primary studies included in the meta-analyses and calculated total pooled ESs and its 95% CIs of the associations between depression and the four peripheral inflammatory biomarkers using a random effects model and SMD as an ES metrics. To assess heterogeneity among primary study ESs, the I^2 index was calculated. We assessed the presence of publication bias using funnel plots and Egger's tests. Data in primary studies presented with SEs were converted as SDs and calculation methods for converting median and IQR to mean and SD were applied in recalculation if necessary.[24,25] If ESs are presented with other metrics rather than SMD, we recalculated SMD with information in primary studies. We adhered to Preferred Reported Items for Systematic Reviews and Meta-analysis (PRISMA) guidelines[26] and registered our review in PROSPERO (CRD42019133888).

Role of the funding source

This review was supported by the Basic Science Research Program through the National

Research Foundation of Korea funded by the Ministry of Science, ICT & Future Planning, Republic of Korea (2017R1A2B3008214). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. All authors had full access to all of the study data, and the corresponding authors had the final responsibility for the decision to submit for publication.

Results

Database

A total of 517 potentially eligible articles were retrieved by the literature search (figure 1). During the screening process, 502 articles were excluded, and 15 meta-analyses were finally included (table 1).[3,4,7-13,20-22,27-29] The publication years of the meta-analyses ranged from 2010 to 2019. All meta-analyses noted significant heterogeneity among primary studies. Only two (13·3%) recently published meta-analyses were registered in PROSPERO.[11,13]

Errors detected across meta-analyses

Errors detected in meta-analyses of the association between depression and the four peripheral inflammatory biomarkers are detailed in table 2. The number of primary studies included in the meta-analyses ranged from 3 to 61. Except for the meta-analysis performed by Rowland et al. (2018), which investigated the association of bipolar depression with TNF-α, IL-6, and CRP,[20] all meta-analyses (93·3%) had at least one type of error. Overall, of the 521 primary studies included in the meta-analyses, errors were identified for 118 (22·6%) studies. The types of errors included incorrect sample sizes (16·9%), incorrectly used SD (29·7%), incorrect participant inclusion (5·9%), calculation error (28·0%), or analysis with insufficient data (19·5%). Among types of error, incorrectly used SD was the most frequent error.

Errors detected across overlapping primary studies

A total of 305 overlapping primary studies were included in the meta-analyses (table 3). Overall, 61 of the 305 primary studies (20·0%) were associated with incorrect data extraction. Data extracted from the overlapping studies of TNF-α, IL-1β, IL-6, and CRP showed errors in 15·2%, 20·5%, 17·5%, and 44·4% of studies, respectively. The types of errors consisted of incorrect

sample sizes (24·6%), incorrectly used SD (44·2%), incorrect participant inclusion (8·2%), calculation error (19·7%), or analysis with insufficient data (3·3%). Again, incorrectly used SD was the most frequent error among overlapping studies.

Re-estimation of meta-analytical findings

Table 4 shows a comparison of originally calculated pooled ESs and their CIs with our recalculated values. Eleven (29·7%) out of the 37 pooled ESs from overlapping meta-analyses changed by more than 0·1 after recalculation and those changes were presented as 'change in results'. In 11 pooled ESs for TNF-α, two (18·2%) recalculated pooled ESs was classified as 'change in results'. Interestingly, five (62·5%) out of eight recalculated pooled ESs for IL-1 β was shown as 'change in results'. In comparison, two pooled ESs for IL-1β by Ellul and colleagues[28] were not included because this meta-analysis did not specify the distinction between high- and low-quality studies. Therefore, we had to recalculate a pooled ES by integrating all studies in the meta-analysis. Although the recalculated pooled ES differed from a non-significant result of low-quality studies, it was unable to determine a change in results because those ESs were derived from non-comparable data. In 12 and 6 pooled ESs for IL-6 and CRP, three (25·0%) and one (16·7%) recalculated pooled ESs were shown as 'change in results', respectively.

We also included all primary studies for each inflammatory biomarker and calculated total pooled SMDs of the associations with depression. The number of primary studies for the four peripheral inflammatory biomarkers ranged from 39 to 112. We found that elevated peripheral levels of three inflammatory biomarkers — TNF-α, IL-6, and CRP — were significantly associated with depression. Total pooled ESs with 95% CIs for TNF-α, IL-6, and CRP were 0·49 (95% CI 0·34, 0·65), 0·46 (95% CI 0·38, 0·54), and 0·27 (95% CI 0·21, 0·33),

respectively. IL-1β was not associated with depression. Significant heterogeneity was found for all four biomarkers, with I^2 values ranging from 85·1% to 88·2% (supplementary figures S1–S4). Funnel plots and Egger's tests showed publication bias among studies of TNF-α, IL-6, and CRP (all $p < 0·001$) but not among studies of IL-1β ($p = 0·257$) (supplementary figures S5–S8).

Discussion

We found a considerable number of errors in 14 (93·3%) of the 15 overlapping meta-analyses of the association between depression and four peripheral inflammatory biomarkers. Of the 521 primary studies included in the overlapping meta-analyses, errors were identified for 118 (22·6%) of them. The most common errors were incorrectly used SD (29·7%) followed by calculation error (28·0%), analysis with insufficient data (19·5%), incorrect sample sizes (16·9%), and incorrect participant inclusion (5·9%). Of 305 overlapping studies that were included in more than one meta-analysis, errors were found in 61 (20·0%) of them. The most common errors were also incorrectly used SD (44·2%) and it followed by incorrect sample sizes (24·6%), calculation error (19·7%), incorrect participant inclusion (8·2%), and analysis with insufficient data (3·1%). After correcting these errors and repeating the analyses, 11 (29·7%) out of 37 pooled ESs from the meta-analyses changed the magnitude level of the effect size. The updated meta-analyses showed that elevated levels of peripheral TNF-α, IL-6, CRP, but not IL-1β, were associated with depression.

Incorrectly identifying the sample sizes was a potential meta-analytical problem. Although this type of error was more prominent among overlapping studies of CRP, it was also noted in studies of the other three inflammatory biomarkers. Because the data extraction process is usually performed manually, it may increase the risk of errors. In future, machine learning may

be applied to searching for and screening studies to include in a meta-analysis and further improve the meta-analytic research.[30]

Incorrectly used SD was the most common data extraction error in our umbrella review, consistent with previous reports of SEs mistaken for SDs.[15,31] This type of error can inflate the point estimate and artificially reduce its CI substantially,[15] impacting the pooled ESs and its estimated heterogeneity. Therefore, it can change the clinical meaningfulness of the meta-analytical results. Some primary studies did not even indicate precisely whether their results were presented with SEs or SDs. In the review of Jones et al. (2005), 34 systematic reviews conducted by the Cochrane Cystic Fibrosis and Genetic Disorders Group were evaluated for data-handling and reporting errors.[16] As a result, errors were found in 20 reviews and 4 (20·0%) out of 20 reviews were related with incorrectly used SD. Thirty-five primary studies in 11 meta-analyses included our study were found with incorrectly used SD. Accordingly, this type of error may be more frequent in meta-analytical studies in psychiatry research than in other medical disciplines.

Inaccuracies in participant inclusion and calculation were also noticed. Some studies which are not related to depression and inflammatory biomarkers or studies that do not meet the inclusion criteria of a meta-analysis were found to be erroneously included in the meta-analytic results.[32,33] Ford et al. (2009) reported in their review that five (62·5%) out of eight meta-analyses of pharmacological interventions for irritable bowel syndrome included studies that were ineligible according to the predefined eligibility criteria.[14] In our study of 15 meta-analyses, only seven primary studies included in three meta-analyses were related to this type of error and the error was relatively infrequent than that of the study conducted by Ford et al..

Cases of calculation error which indicate inaccurate effect sizes despite no errors found in reported primary study data lead us to presume a possible error of discrepancy between used

data for actual calculation and reported data.[34] In the review of Gøtzsche et al. (2007), 27 meta-analyses that had used the SMD and were published in 2004 were included for the evaluation of errors. The authors randomly selected two trials from each meta-analysis and found that 10 (37%) of the 27 meta-analyses has at least one error. In the ten meta-analyses with errors, one (8·3%) out of 12 trials was related to calculation errors. Although it is challenging to compare the calculation error rate of our results to that of the review directly, we can presume that considerable calculation error may also influence results in psychiatry discipline.

In some cases, primary studies did not report sufficient information about their analyses. Therefore, it was not possible to extract data from some primary studies,[35,36] because essential data for meta-analysis were missing. In line with this, another problem that nonstatistically significant effects (NSUEs) are frequently unreported should be addressed. Some studies with nonsignificant group differences sometimes did not present any statistics that are necessary to be converted into ES. Recent statistical approaches (e.g., MetaNSUE) have been developed to overcome this problem.[37]

Like many previous meta-analyses,[4,8-10,12,21] total pooled ESs of the four peripheral inflammatory biomarkers in our study showed that elevated peripheral levels of TNF-α, IL-6, and CRP, but not IL-1β, were associated with depression. We also found significant heterogeneity among primary studies of all four biomarkers, presumably reflecting diversity in the characteristics of individual studies and the important roles of biological, clinical, and technical confounders.

Some limitations of our umbrella review and meta-analysis should be acknowledged. Severity and duration of depression, medication status, and other confounding factors such as body mass index were not fully adjusted. Significant heterogeneity among primary studies also makes the interpretation for the total pooled ESs of the four biomarkers limited. We included unipolar and

bipolar depression and summarized those data together. As differences between unipolar and bipolar depression are reported,[38] this should be taken into consideration before generalizing our summary results. In addition, data were extracted from graphs in 51 primary studies. Data extracted from graphs may be less accurate than data extracted from numbers and incorrectly used SD and calculation error can be related to inaccuracies of data extraction from graphs. However, only 3 cases of incorrectly used SD and 9 cases of calculation error were noticed in all extracted data from graphs and our results were not primarily affected by it. Lastly, although we spent much time and effort checking for the presence of errors in previous meta-analyses, the possibility of errors in our umbrella review itself cannot be excluded.

Although the statistical calculations in meta-analyses are ostensibly simple, data extraction and analysis are particularly prone to errors. The high prevalence of errors that can negate or even reverse the significance of findings. Because errors in data extraction may influence ES and inflate heterogeneity among studies, efforts to reduce data extraction errors are important in such studies for the association between depression and peripheral inflammatory biomarkers where high heterogeneity and conflicting results have been reported continuously.

Contributors

SL, KHL, EL, and JIS designed the study. SL, KMP, SJP, and WJK performed the literature search and screening; extracted, analyzed, and interpreted the data; and made the figures and tables; any discrepancies were resolved via discussion among SL, KHL, EL, and JIS. SL, EL, and JIS drafted the manuscript. JL, AK, LS, MS, BS, AK (Koyanagi), LJ, AS, TT, ED, HO, ARB, AFC, JR, SKA, KN, and PFP were involved in critically revising the manuscript for important intellectual content. All authors approved the final version of the manuscript for publication. EL and JIS contributed as joint corresponding authors.

Declaration of interests

We declare no competing interests.

Acknowledgments

This paper presents independent research. The views expressed in this publication are those of the authors and not necessarily those of the acknowledged institutions.

Research in context

Evidence before this study

We searched PubMed/MEDLINE, Embase, PsycINFO, and the Cochrane Library for meta-analyses of observational studies of the association between depression and four peripheral inflammatory biomarkers (TNF-α, IL-1β, IL-6, and CRP) published between database inception and January 14, 2020, without language restrictions using the following search terms: 'CRP OR IL-1beta OR IL-6 OR TNF-alpha', 'depress*', and 'meta'. We identified several meta-analyses that report conflicting results. For instance, a meta-analysis of studies conducted among older adults reports that depression is associated with IL-1β and IL-6 but not TNF-α and CRP. However, other meta-analyses conducted among general populations report that unipolar and bipolar depression are associated with TNF-α and CRP. Errors in data extraction could be one of the reasons why different meta-analyses reach different conclusions for the same research question.

Added value of this study

To our knowledge, this review is the first to evaluate the presence, frequency, and nature of errors and their impact on meta-analytic results in psychiatry research. Of the 15 meta-analyses included, we found a substantial number of errors in 14 meta-analyses, including incorrect sample sizes, incorrectly used SD, incorrect participant inclusion, calculation error, and analysis with insufficient data. After correcting the errors and performing recalculations, 11 out

of 37 pooled ESs from the meta-analyses changed the magnitude level of the effect size more than 0·1. In addition, our total pooled ESs from all primary studies indicate that elevated peripheral levels of TNF-α, IL-6, and CRP, but not IL-1β, are associated with depression.

Implications of all the available evidence

Our findings suggest that data extraction in meta-analysis is particularly prone to errors. As errors can negate or even reverse significant findings in meta-analyses, reducing data extraction errors is important in studies for the association between depression and peripheral inflammatory biomarkers where high heterogeneity and conflicting results have been reported continuously. In addition, the association between three peripheral inflammatory biomarkers and depression found in our total pooled ESs provides further evidence that alterations in immune-inflammatory pathways play important roles in the pathophysiology of depression.

References

1 Maes M. Evidence for an immune response in major depression: a review and hypothesis. *Prog Neuropsychopharmacol Biol Psychiatry* 1995; **19**: 11-38.

2 Miller AH, Raison CL. The role of inflammation in depression: from evolutionary imperative to modern treatment target. *Nat Rev Immunol* 2016; **16**: 22-34.

3 Howren MB, Lamkin DM, Suls J. Associations of depression with C-reactive protein, IL-1, and IL-6: a meta-analysis. *Psychosom Med* 2009; **71**: 171-86.

4 Liu Y, Ho RC, Mak A. Interleukin (IL)-6, tumour necrosis factor alpha (TNF-alpha) and soluble interleukin-2 receptors (sIL-2R) are elevated in patients with major depressive disorder: a meta-analysis and meta-regression. *J Affect Disord* 2012; **139**: 230-9.

5 Wium-Andersen MK, Orsted DD, Nielsen SF, Nordestgaard BG. Elevated C-reactive protein levels, psychological distress, and depression in 73, 131 individuals. *JAMA Psychiatry* 2013; **70**: 176-84.

6 Berlin JA, Golub RM. Meta-analysis as Evidence: Building a Better Pyramid. *JAMA* 2014; **312**: 603-6.

7 Ng A, Tam WW, Zhang MW, et al. IL-1beta, IL-6, TNF- alpha and CRP in Elderly Patients with Depression or Alzheimer's disease: Systematic Review and Meta-Analysis. *Sci Rep* 2018; **8**: 12050.

8 Kohler CA, Freitas TH, Maes M, et al. Peripheral cytokine and chemokine alterations in depression: a meta-analysis of 82 studies. *Acta Psychiatr Scand* 2017; **135**: 373-87.

9 Goldsmith DR, Rapaport MH, Miller BJ. A meta-analysis of blood cytokine network alterations in psychiatric patients: comparisons between schizophrenia, bipolar disorder and depression. *Mol Psychiatry* 2016; **21**: 1696-709.

10 Dowlati Y, Herrmann N, Swardfager W, et al. A meta-analysis of cytokines in major depression. *Biol Psychiatry* 2010; **67**: 446-57.

11 D'Acunto G, Nageye F, Zhang J, Masi G, Cortese S. Inflammatory Cytokines in Children and Adolescents with Depressive Disorders: A Systematic Review and Meta-Analysis. *J Child Adolesc Psychopharmacol* 2019; **29**: 362-9.

12 Osimo EF, Baxter LJ, Lewis G, Jones PB, Khandaker GM. Prevalence of low-grade inflammation in depression: a systematic review and meta-analysis of CRP levels. *Psychol Med* 2019; **49**: 1958-70.

13 Perrin AJ, Horowitz MA, Roelofs J, Zunszain PA, Pariante CM. Glucocorticoid Resistance: Is It a Requisite for Increased Cytokine Production in Depression? A Systematic Review and Meta-Analysis. *Front Psychiatry* 2019; **10**: 423.

14 Ford AC, Guyatt GH, Talley NJ, Moayyedi P. Errors in the conduct of systematic reviews of pharmacological interventions for irritable bowel syndrome. *Am J Gastroenterol* 2010; **105**: 280-8.

15 Gotzsche PC, Hrobjartsson A, Maric K, Tendal B. Data extraction errors in meta-analyses that use standardized mean differences. *Jama* 2007; **298**: 430-7.

16 Jones AP, Remmington T, Williamson PR, Ashby D, Smyth RL. High prevalence but low impact of data extraction and reporting errors were found in Cochrane systematic reviews. *J Clin Epidemiol* 2005; **58**: 741-2.

17 Park JH, Eisenhut M, van der Vliet HJ, Shin JI. Statistical controversies in clinical research: overlap and errors in the meta-analyses of microRNA genetic association studies in cancers. *Ann Oncol* 2017; **28**: 1169-82.

18 Messori A, Scroccaro G, Martini N. Calculation errors in meta-analysis. *Ann Intern Med* 1993; **118**: 77-8.

19 Buscemi N, Hartling L, Vandermeer B, Tjosvold L, Klassen TP. Single data extraction generated more errors than double data extraction in systematic reviews. *Journal of clinical epidemiology* 2006; **59**: 697-703.

20 Rowland T, Perry BI, Upthegrove R, et al. Neurotrophins, cytokines, oxidative stress mediators and mood state in bipolar disorder: systematic review and meta-analyses. *Br J Psychiatry* 2018; **213**: 514-25.

21 Fernandes BS, Steiner J, Molendijk ML, et al. C-reactive protein concentrations across the mood spectrum in bipolar disorder: a systematic review and meta-analysis. *Lancet Psychiatry* 2016; **3**: 1147-56.

22 Dargél AA, Godin O, Kapczinski F, Kupfer DJ, Leboyer M. C-reactive protein alterations in bipolar disorder: A meta-analysis. *The Journal of Clinical Psychiatry* 2015; **76**: 142-50.

23 GetData Graph Digitizer, 2013. http://www.getdata-graph-digitizer.com (accessed 18 November, 2019.

24 Luo D, Wan X, Liu J, Tong T. Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. *Stat Methods Med Res* 2018; **27**: 1785-805.

25 Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC medical research methodology* 2014; **14**: 135.

26 Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine* 2009; **151**: 264-9.

27 Bizik G. Meta-analysis of plasma interleukine-6 levels in patients with depressive

disorder. *Act Nerv Super* 2010; **52**: 76-80.

28 Ellul P, Boyer L, Groc L, Leboyer M, Fond G. Interleukin-1 beta-targeted treatment strategies in inflammatory depression: toward personalized care. *Acta Psychiatr Scand* 2016; **134**: 469-84.

29 Munkholm K, Vinberg M, Vedel Kessing L. Cytokines in bipolar disorder: a systematic review and meta-analysis. *J Affect Disord* 2013; **144**: 16-27.

30 Xiong Z, Liu T, Tse G, et al. A Machine Learning Aided Systematic Review and Meta-Analysis of the Relative Risk of Atrial Fibrillation in Patients With Diabetes Mellitus. *Front Physiol* 2018; **9**: 835.

31 Tendal B, Higgins JP, Juni P, et al. Disagreements in meta-analyses using outcomes measured on continuous or rating scales: observer agreement study. *Bmj* 2009; **339**: b3128.

32 Koenig HG, Cohen HJ, George LK, Hays JC, Larson DB, Blazer DG. Attendance at religious services, interleukin-6, and other biological parameters of immune function in older adults. *Int J Psychiatry Med* 1997; **27**: 233-50.

33 Komulainen P, Lakka TA, Kivipelto M, et al. Serum high sensitivity C-reactive protein and cognitive function in elderly women. *Age Ageing* 2007; **36**: 443-8.

34 Soneji S. Errors in Data Input in Meta-analysis on Association Between Initial Use of e-Cigarettes and Subsequent Cigarette Smoking Among Adolescents and Young Adults. *JAMA Pediatr* 2018; **172**: 92-3.

35 Steptoe A, Kunz-Ebrecht SR, Owen N. Lack of association between depressive symptoms and markers of immune and vascular inflammation in middle-aged men and women. *Psychol Med* 2003; **33**: 667-74.

36 Suarez EC. C-reactive protein is associated with psychological risk factors of

cardiovascular disease in apparently healthy adults. *Psychosom Med* 2004; **66**: 684-91.

37 Radua J, Schmidt A, Borgwardt S, et al. Ventral Striatal Activation During Reward Processing in Psychosis: A Neurofunctional Meta-Analysis. *JAMA psychiatry* 2015; **72**: 1243-51.

38 Goya-Maldonado R, Brodmann K, Keil M, Trost S, Dechent P, Gruber O. Differentiating unipolar and bipolar depression by alterations in large-scale brain networks. *Human brain mapping* 2016; **37**: 808-18.

Figure Legends

Figure 1. Flow chart of literature search and screening.

Table 1. Literature search, analysis, and reporting of overlapping meta-analyses of the association between depression and inflammatory biomarkers.

First author, Year	Publication date	Search date	Included analyses				PROSPERO Registration
			Effect size metrics	Heterogeneity	Publication bias	Meta-regression	
TNF-α							
D'Acunto, 2019[11]	June 2019	July 2018	Hedges' *g*	+	Egger's test	-	+
Perrin, 2019[13]	June 2019	October 2018	SMD	+	Funnel plot	+	-
Ng, 2018[7]	August 2018	March 2017	SMD	+	Egger's test	+	-
Köhler, 2017[8]	January 2017	May 2016	Hedges' *g*	+	Funnel plot, Egger's test, trim and fill, fail-safe N	+	-
Goldsmith, 2016[9]	February 2016	March 2015	Hedges' *g*	+	Funnel plot, Egger's test	+	-
Liu, 2012[4]	August 2012	February 2011	SMD	+	Egger's test	+	-
Dowlati, 2010[10]	March 2010	August 2009	MD	+	Funnel plots, rank correlation tests	-	N/A
Rowland, 2018[20]	September 2018	February 2017	SMD	+	Funnel plot	-	-
Munkholm, 2013[29]	January 2013	January 2012	SMD	+	No	-	-
IL-1β							
D'Acunto, 2019[11]	June 2019	July 2018	Hedges' *g*	+	Egger's test	-	+
Ng, 2018[7]	August 2018	March 2017	SMD	+	Egger's test	+	-
Köhler, 2017[8]	January 2017	May 2016	Hedges' *g*	+	Funnel plot, Egger's test, trim and fill, fail-safe N	+	-
Goldsmith, 2016[9]	February 2016	March 2015	Hedges' *g*	+	Funnel plot, Egger's test	+	-
Liu, 2012[4]	August 2012	February 2011	SMD	+	Egger's test	+	-
Dowlati, 2010[10]	March 2010	August 2009	MD	+	Funnel plots, rank correlation tests	-	N/A
Howren, 2009[3]	February 2009	January 2008	SMD	+	Funnel plot, fail-safe N	+	N/A
Ellul, 2016[28]	December 2016	January 2016	SMD	+	Funnel plot, Egger's test	-	-
IL-6							
Perrin, 2019[13]	June 2019	October 2018	SMD	+	Funnel plot	+	-
Ng, 2018[7]	August 2018	March 2017	SMD	+	Egger's test	+	-
Köhler, 2017[8]	January 2017	May 2016	Hedges' *g*	+	Funnel plot, Egger's test, trim and fill, fail-safe N	+	-
Goldsmith, 2016[9]	February 2016	March 2015	Hedges' *g*	+	Funnel plot, Egger's test	+	-
Liu, 2012[4]	August 2012	February 2011	SMD	+	Egger's test	+	-
Bizik, 2010[27]	June 2010	October 2009	SMD	+	Fail-safe N	-	N/A
Dowlati, 2010[10]	March 2010	August 2009	MD	+	Funnel plots, rank correlation tests	-	N/A
Howren, 2009[3]	February 2009	January 2008	SMD	+	Funnel plot, fail-safe N	+	N/A
Rowland, 2018[20]	September 2018	February 2017	SMD	+	Funnel plot	-	-
Munkholm, 2013[29]	January 2013	January 2012	SMD	+	No	-	-
CRP							
Osimo, 2019[12]	September 2019	July 2018	OR	+	Funnel plot, Egger's test	+	+
Ng, 2018[7]	August 2018	March 2017	SMD	+	Egger's test	+	-
Howren, 2009[3]	February 2009	January 2008	SMD	+	Funnel plot, fail-safe N	+	N/A
Rowland, 2018[20]	September 2018	February 2017	SMD	+	Funnel plot	-	-
Fernandes, 2016[21]	December 2016	August 2016	Hedges' *g*	+	Funnel plot, trim and fill, fail-safe N test	+	-
Dargel, 2015[22]	February 2015	June 2013	SMD	+	Funnel plot, Egger's test	-	-

PROSPERO, international prospective register for systematic review protocols; SMD, standardized mean difference; MD, mean difference; OR, odds ratio TNF-α, tumor necrosis factor-α; IL-1β, interleukin 1-β; IL-6, interleukin 6; CRP, C-reactive protein; N/A, not applicable.

Table 2. Results of overlapping meta-analyses of the association between depression and inflammatory biomarkers.

First author, Year	Included diagnosis of patients	No. of studies/papers used	No. of cases/controls	Statistical significance	Heterogeneity	Publication bias	Errors				
							Incorrect sample sizes	Incorrectly used SD	Incorrect participant inclusion	Calculation error	Analysis with insufficient data[a]
TNF-α											
D'Acunto, 2019[11]	DD	4	72/61	No	No	Yes	0	0	0	0	0
Perrin, 2019[13]	DD	15	604/864	Yes	Yes	N/A	0	2	0	0	0
Ng, 2018[7]	DD	5	478/4611	No	Yes	N/A	0	0	0	0	0
Köhler, 2017[8]	MDD	42	1742/1478	Yes	Yes	Yes	2	1	1	2	0
Goldsmith, 2016[9]	MDD, BD	BD: 2 Acute MDD: 8 Chronic MDD: 8	BD: 44/105 Acute: 296/281 Chronic: 362/439	BD: No Acute MDD: Yes Chronic MDD: No	BD: No Acute MDD: Yes Chronic MDD: Yes	BD: N/A Acute MDD: No Chronic MDD: N/A	0	Chronic MDD: 1	0	0	0
Liu, 2012[4]	MDD	15	541/444	Yes	Yes	No	0	1	0	1	0
Dowlati, 2010[10]	MDD	13	438/350	Yes	Yes	No	0	4	0	0	0
Rowland, 2018[20]	BD	6	81/253	Yes	Yes	N/A	0	0	0	0	0
Munkholm, 2013[29]	BD	3	39/155	No	Yes	N/A	0	2	0	0	0
IL-1β											
D'Acunto, 2019[11]	DD	4	72/61	No	Yes	No	0	1	0	0	0
Ng, 2018	DD	5	314/895	Yes	Yes	N/A	0	0	0	0	0
Köhler, 2017[8]	MDD	22	784/722	No	No	No	1	0	0	0	0
Goldsmith, 2016[9]	MDD, BD	Acute MDD: 4 Chronic MDD: 4	Acute MDD: 116/112 Chronic MDD: 138/190	Acute MDD: Yes Chronic MDD: No	Acute MDD: Yes Chronic MDD: Yes	N/A	0	Chronic MDD: 1	0	0	0
Ellul, 2016[28]	MDD	21	824/1085	High-quality studies: Yes Low-quality studies: No	High-quality studies: Yes Low-quality studies: Yes	N/A	3	2	0	0	2
Liu, 2012[4]	MDD	10	290/290	No	Yes	No	1	1	0	0	0
Dowlati, 2010[10]	MDD	9	267/246	No	Yes	No	1	2	0	0	0
Howren, 2009[3]	DD	14	323/346[b]	Yes	Yes	Yes	0	1	1	4	1
IL-6											
Perrin, 2019[13]	DD	23	1607/1042	Yes	Yes	N/A	0	3	0	0	0
Ng, 2018[7]	DD	9	2016/7211	Yes	Yes	No	0	0	0	1	0
Köhler, 2017[8]	MDD	42	1587/1183	Yes	No	No	0	1	1	0	0
Goldsmith, 2016[9]	MDD, BD	BD: 3 Acute MDD: 10 Chronic MDD: 7	BD:102/344 Acute MDD: 306/216 Chronic MDD: 189/211	BD: No Acute MDD: Yes Chronic MDD: Yes	BD: Yes Acute MDD: Yes Chronic MDD: Yes	BD: N/A Acute MDD: No Chronic MDD: No	0	Chronic MDD: 2	0	0	0
Liu, 2012[4]	MDD	18	508/415	Yes	Yes	No	1	2	0	0	0
Bizik, 2010[27]	DD	16	433/581	Yes	Yes	Yes	1	3	0	0	0
Dowlati, 2010[10]	MDD	16	492/400	Yes	Yes	No	2	4	0	0	0
Howren, 2009[3]	DD	61	3020/10598[b]	Yes	Yes	Yes	1	0	2	12	12
Rowland, 2018[20]	BD	6	130/471	No	Yes	N/A	0	0	0	0	0
Munkholm, 2013[29]	BD	3	88/352	No	Yes	N/A	0	0	0	0	0
CRP											
Osimo, 2019[12]	DD	17	7761/155728	Yes	Yes	No	2	0	1	0	0
Ng, 2018[7]	DD	9	2513/11991	No	Yes	No	0	1	0	1	0
Howren, 2009[3]	DD	49	4050/23179[b]	Yes	Yes	Yes	1	0	1	8	8
Rowland, 2018[20]	BD	3	91/329	No	No	N/A	0	0	0	0	0
Fernandes, 2016[21]	BD	11	441/922	Yes	Yes	No	2	0	0	4	0
Dargel, 2015[22]	BD	4	107/297	No	Yes	N/A	2	0	0	0	0
Total											
Primary studies		521					20	35	7	33	23
(Meta-analyses)		(15)					(9)	(11)	(3)	(5)	(2)

SMD, standardized mean difference; DD, depressive disorder, BD, bipolar depression; MDD, major depressive disorder; TNF-α, tumor necrosis factor-α; IL-1β, interleukin 1-β; IL-6, interleukin 6; CRP, C-reactive protein; N/A, not applicable.

[a]Analysis with insufficient data indicates that sufficient information to calculate SMD was not provided in a primary study.

[b]Effect sizes in some single population studies were calculated using correlation coefficient. In such cases, the number of all participants in the study is not included in this column.

Table 3. Summary characteristics of overlapping primary studies of inflammatory biomarkers with errors.

Characteristic of error	Number of studies (%)	
TNF-α		
Total No. of overlapping studies of TNF-α	79 (100%)	
Total No. of overlapping studies of TNF-α with errors	**12 (15·2%)**	
Incorrect sample sizes		1 (8·3%)
Incorrectly used SD		8 (66·7%)
Incorrect participant inclusion		1 (8·3%)
Calculation error		2 (16·7%)
Analysis with insufficient data		0 (0%)
IL-1β		
Total No. of overlapping studies of IL-1β	73 (100%)	
Total No. of overlapping studies of IL-1β with errors	**15 (20·5%)**	
Incorrect sample sizes		4 (26·7%)
Incorrectly used SD		6 (40·0%)
Incorrect participant inclusion		1 (6·7%)
Calculation error		2 (13·3%)
Analysis with insufficient data		2 (13·3%)
IL-6		
Total No. of overlapping studies of IL-6	126 (100%)	
Total No. of overlapping studies of IL-6 with errors	**22 (17·5%)**	
Incorrect sample sizes		5 (22·7%)
Incorrectly used SD		12 (54·5%)
Incorrect participant inclusion		1 (4·6%)
Calculation error		4 (18·2%)
Analysis with insufficient data		0 (0%)
CRP		
Total No. of overlapping studies of CRP	27 (100%)	
Total No. of overlapping studies of CRP with errors	**12 (44·4%)**	
Incorrect sample sizes		5 (41·7%)
Incorrectly used SD		1 (8·3%)
Incorrect participant inclusion		2 (16·7%)
Calculation error		4 (33·3%)
Analysis with insufficient data		0 (0%)
Total for all four peripheral inflammatory biomarkers		
Total No. of overlapping studies of all peripheral inflammatory biomarkers	305 (100%)	
Total No. of overlapping studies of all peripheral inflammatory biomarkers with errors	**61 (20·0%)**	
Incorrect sample sizes		15 (24·6%)
Incorrectly used SD		27 (44·2%)
Incorrect participant inclusion		5 (8·2%)
Calculation error		12 (19·7%)
Analysis with insufficient data		2 (3·3%)

SD, standard deviation; TNF-α, tumor necrosis factor-α; IL-1β, interleukin 1-β; IL-6, interleukin 6; CRP, C-reactive protein. 'Analysis with insufficient data' indicates that sufficient information to calculate SMD was not provided in a primary study.

Table 4. Comparison of results of overlapping meta-analyses with recalculated effect sizes and confidence intervals.

First author, year	Model	Outcome metrics	Reported		Recalculated		Change in results[a]	Type of data extraction error
			ES (95% CI)	p-value	ES (95% CI)	p-value		
TNF-α								
D'Acunto, 2019[11]	Random effects	Hedges' *g*	0·35 (-0·01, 0·70)	0·053	**0·35 (0·01, 0·70)**	**0·045**	None	No errors.
Perrin, 2019[13]	Random effects	SMD	0·40 (0·12, 0·68)	0·006	0·49 (0·16, 0·82)	0·004	None	SD (Carvalho, 2013; Lamers, 2013).
Ng, 2018[7]	Random effects	SMD	0·11 (-0·12, 0·35)	0·351	0·12 (-0·12, 0·35)	0·334	None	No errors.
Köhler, 2017[8]	Random effects	Hedges' *g*	0·68 (0·43, 0·92)	<0·001	0·65 (0·41, 0·88)	<0·001	None	Sample (Eller, 2009; Farid Hosseini, 2007), SD (O'Donovan, 2013), PI (O'Brien, 2007), and calculation (Baek, 2013; Grassi-Oliveira, 2009),
Goldsmith, 2016[9]	Fixed effect	Hedges' *g*	-0·16 (-0·53, 0·21)	0·4	-0·16 (-0·53, 0·21)	0·386	None	No errors in the BD subgroup.
			0·35 (0·17, 0·53)	<0·01	0·39 (0·20, 0·57)	<0·001	None	No errors in the acute MDD subgroup.
			0·05 (-0·10, 0·19)	0·52	0·07 (-0·08, 0·22)	0·332	None	SD (Einvik, 2012) in the chronic MDD subgroup.
Liu, 2012[4]	Random effects	SMD	0·56 (0·13, 0·99)	0·01	0·49 (0·10, 0·89)	0·014	None	SD (Pavon, 2006) and calculation (Tuglu, 2003).
Dowlati, 2010[10]	Random effects	MD	3·97 (2·24, 5·71)	<0·001	**3·14 (1·61, 4·66)**	**<0·001**	**Changed[b]**	SD (O'Brien, 2007; Pavon, 2006; Tuglu, 2003; Yang, 2007).
Rowland, 2018[20]	Random effects	SMD	2·09 (0·82, 3·36)	<0·001	2·17 (0·85, 3·49)	0·001	None	No errors.
Munkholm, 2013[29]	Random effects	SMD	4·31 (-0·57, 9·19)	0·08	**3·16 (0·02, 6·29)**	**0·048**	**Changed**	SD (Kapczinski, 2011; O'Brien, 2006).
Total pooled results of all primary studies	Random effects	SMD	-	-	**0·49 (0·34, 0·65)**	**<0·001**	-	71 primary studies included.
IL-1β								
D'Acunto, 2019[11]	Random effects	Hedges' *g*	0·47 (-0·20, 1·15)	0·169	**0·26 (-0·27, 0·78)**	**0·335**	**Changed**	SD (Miklowitz, 2016).
Ng, 2018[7]	Random effects	SMD	0·64 (0·06, 1·21)	0·026	0·65 (0·08, 1·22)	0·026	None	No errors.
Köhler, 2017[8]	Random effects	Hedges' *g*	0·03 (-0·29, 0·35)	0·847	**0·16 (-0·21, 0·53)**	**0·402**	**Changed**	Sample (Alcocer-Gomez, 2014).
Goldsmith, 2016[9]	Fixed effect	Hedges' *g*	-0·22 (-0·49, 0·06)	0·13	-0·19 (-0·46, 0·08)	0·164	None	No errors in the acute MDD subgroup.
			0·21 (-0·04, 0·47)	0·1	0·22 (-0·04, 0·47)	0·096	None	SD (Einvik, 2012) in the chronic MDD subgroup.
Ellul, 2016[28]	Random effects	SMD	-0·54 (-1·03, -0·83) [c]	0·021	**0·36 (0·03, 0·70)**	**0·035**	-	Sample (Alcocer-Gomez, 2013; Marques-Deak, 2007; van den Biggelaar, 2006), SD (Pavon, 2006; Piletz, 2006), and ID (Hughes, 2012; Yang, 2007).
			0·10 (-0·45, 0·66)	0·715				
Liu, 2012[4]	Random effects	SMD	-0·53 (-1·36, 0·32)	0·221	**-0·13 (-0·79, 0·53)**	**0·697**	**Changed**	Sample (Kagaya, 2001) and SD (Pavon, 2006).
Dowlati, 2010[10]	Random effects	MD	-1·58 (-3·59, 0·43)	0·39	**-1·35 (-3·69, 1·00)**	**0·26**	**Changed[b]**	Sample (Kagaya, 2001) and SD (Pavon, 2006; Yang, 2007).
Howren, 2009[3]	Random effects	SMD	0·35 (0·03, 0·67)	0.03	**0·24 (-0·15, 0·63)**	**0·229**	**Changed**	SD (Kagaya, 2001), PI (Levine, 1999), calculation (Ferketich, 2005; Huang, 2007; Miller, 2002; Owen, 2001), and ID (Hekler, 2007).
Total pooled results of all primary studies	Random effects	SMD	-	-	0·17 (-0·06, 0·39)	0·143	-	39 primary studies included.
IL-6								
Perrin, 2019[13]	Random effects	SMD	0·61 (0·36, 0·85)	<0·001	**0·50 (0·27, 0·74)**	**<0·001**	**Changed**	SD(Carvalho, 2013; Lamers, 2013, Maes, 1995c).
Ng, 2018[7]	Random effects	SMD	0·38 (0·16, 0·60)	<0·001	0·38 (0·16, 0·60)	<0·001	None	Calculation (Nadroski, 2016).
Köhler, 2017[8]	Random effects	Hedges' *g*	0·62 (0·49, 0·76)	<0·001	0·64 (0·50, 0·78)	<0·001	None	SD (O'Donovan, 2013) and PI (O'Brien, 2007).
Goldsmith, 2016[9]	Fixed effect	Hedges' *g*	0 (-0·23, 0·23)	0·98	0 (-0·23, 0·23)	0·99	None	No errors in the BD subgroup.
			0·76 (0·56, 0·95)	<0·01	0·74 (0·55, 0·92)	<0·001	None	No errors in the acute MDD subgroup.
			0·39 (0·2, 0·59)	<0·01	0·40 (0·20, 0·60)	<0·001	None	SD (Dhabhar, 2009; Einvik, 2012) in the chronic MDD subgroup.
Liu, 2012[4]	Random effects	SMD	0·68 (0·44, 0·92)	<0·001	0·61 (0·38, 0·84)	<0·001	None	Sample (Kagaya, 2001) and SD (Dhabhar, 2009; Pavon, 2006).
Bizik, 2010[27]	Random effects	SMD	1·06 (0·59, 1·52)	<0·001	**0·71 (0·43, 0·99)**	**<0·001**	**Changed**	Sample (Kagaya, 2001) and SD (Alesci, 2005; Dhabhar, 2009; Maes, 1995a).
Dowlati, 2010[10]	Random effects	MD	1·78 (1·23, 2·33)	<0·001	**1·87 (0·92, 2·81)**	**<0·001**	**Changed[b]**	Sample (Kagaya, 2001; O'Brien, 2007) and SD (Dhabhar, 2009; Maes, 1995a; Pavon, 2006; Yang, 2007).
Howren, 2009[3]	Random effects	SMD	0·25 (0·18, 0·31)	<0·001	0·27 (0·19, 0·34)	<0·001	None	Sample (Kagaya, 2001), PI (Cyranowski, 2007; Lutgendorf, 1999), calculation (Ferketich, 2005; Jacobson, 2008; Jehn, 2006; Kiecolt-Glaser, 2007; Kudoh, 2001; Maes, 1995a; Maes, 1997; Miller, 2002; Motivala, 2005; Sluzewska, 1995; Soygur, 2007 (cancer and normal)), and ID (Allen-Mersh, 1998; Costanzo, 2005; Ferruci, 2002; Glaser, 2003; Haack, 1999; Hekler, 2007; Koening, 1997; Ranjit, 2007; Steptoe, 2003; Suarez, 2003; Whooley, 2007 (males and females)).

Rowland, 2018[20]	Random effects	SMD	0·67 (-0·08, 1·42)	0·08	0·63 (-0·13, 1·39)	0·106	None	No errors.
Munkholm et al., 2013[29]	Random effects	SMD	1·04 (-0·54, 2·62)	0·2	1·05 (-0·45, 2·55)	0·17	None	No errors.
Total pooled results of all primary studies	Random effects	SMD	-	-	**0·46 (0·38, 0·54)**	**<0·001**	-	112 primary studies included.
CRP								
Osimo, 2019[12]	Random effects	OR	1·46 (1·22, 1·75)	<0·001	1·40 (1·31, 1·50)	<0·001	None	Sample (Cepeda, 2016; Ekinci, 2017) and PI (Kling, 2007).
Ng, 2018[7]	Random effects	SMD	0.5 (0, 1)	0·05	**0·19 (-0·01, 0·39)**	**0·062**	**Changed**	SD (Kop, 2002) and calculation (Bremmer, 2008).
Howren, 2009[3]	Random effects	SMD	0·22 (0·15, 0·28)	<0·001	0·14 (0·09, 0·19)	<0·001	None	Sample (Almeida, 2007), PI (Kling, 2006), calculation (Arai, 2006; Hornig, 1998; Hung, 2007; Liukkonen, 2006 (males and females); Miller, 2002; Shimbo, 2006; Vaccarino, 2007), and ID (Danner, 2003 (males and females); Douglas, 2004 (males and females); Komulainen, 2007; Ranjit, 2007; Steptoe, 2003; Suares, 2004).
Rowland, 2018[20]	Random effects	SMD	-0·02 (-0·25, 0·21)	0·86	-0·02 (-0·25, 0·21)	0·86	None	No errors.
Fernandes, 2016[21]	Random effects	SMD	0·67 (0·23, 1·11)	0·003	0·74 (0·32, 1·16)	0·001	None	Sample (Cunha, 2008; Jacoby, 2016) and calculation (Bai, 2013; Dickerson, 2015; Hung, 2007; Su, 2011).
Dargel, 2015[22]	Random effects	SMD	0·28 (-0·17, 0·73)	0·227	0·31 (-0·17, 0·78)	0·206	None	Sample (Cunha, 2008; Fontoura, 2012).
Total pooled results of all primary studies	Random effects	SMD	-	-	**0·27 (0·21, 0·33)**	**<0·001**	-	80 primary studies included.

BD, bipolar depression; MDD, major depressive disorder; SMD, standardized mean difference; MD, mean difference; ES, effect size; SD, standard deviation; CI, confidence interval; PI, participant inclusion; ID, insufficient data; TNF-α, tumor necrosis factor-α; IL-1β, interleukin 1-β; IL-6, interleukin 6; CRP, C-reactive protein.

[a]If our recalculation of the pooled ES for each meta-analysis differed from that of the original meta-analysis by 0·1 or more, this is denoted as 'Changed'.

[b]When the MD was converted to SMD, the difference in pooled SMD between the original meta-analysis and our recalculation was 0·1 or more.

[c]The first row of reported ESs presents the results of high-quality studies, and the second row presents the result of low-quality studies. The recalculated ES results from all primary studies.