Dalili, M. N., Penton-Voak, I. S., Harmer, C., & Munafo, M. R. (2015). Meta-analysis of emotion recognition deficits in major depressive disorder. Psychological Medicine, 45(06), 1135–1144. DOI: 10.1017/S0033291714002591

Peer reviewed version

Running Head: Major Depression and Emotion Recognition

Meta-Analysis of Emotion Recognition Deficits in Major Depressive Disorder.

Michael N. Dalili [1,2], Ian S. Penton-Voak [1], Catherine J. Harmer [3], Marcus R. Munafò [1,2,4]

1. School of Experimental Psychology, University of Bristol, UK.
2. MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, UK.
3. Department of Psychiatry, University of Oxford, UK.
4. UK Centre for Tobacco Control Studies, University of Bristol, UK.

Corresponding Author: Michael N. Dalili, School of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol, United Kingdom. T: +44.117. 3310494; F: +44.117.9288588; E: michael.dalili@bristol.ac.uk

Abstract

Background. Many studies have explored associations between depression and facial emotion recognition. However, these studies have used various paradigms and multiple stimulus sets, rendering comparisons difficult. Few studies have attempted to determine the magnitude of any effect and whether studies are properly powered to detect it. We conducted a meta-analysis to synthesize the findings across studies on emotion recognition in depressed individuals compared to controls.

Method. Studies of emotion recognition that included depressed and control samples and published before June 2013 were identified in PubMed and Web of Science. Studies using schematic faces, neuroimaging studies and drug treatment studies were excluded.

Results. Meta-analysis of $k = 22$ independent samples indicated impaired recognition of emotion ($k = 22$, g = -0.16, 95% CI -0.25 to -0.07, $p < 0.001$). Critically, this was observed for anger, disgust, fear, happiness and surprise (ks = 7 to 22, gs = -0.42 to -0.17, ps < 0.08), but not sadness ($k = 21$, $g = -0.09$, 95% CI -0.23 to +0.06, $p = 0.23$). Study level characteristics did not appear to be associated with the observed effect. Power analysis indicated that a sample of approximately 615 cases and 615 controls would be required to detect this association with 80% power at an alpha level of 0.05.

Conclusions. These findings suggest that the emotion recognition impairment reported in the depression literature exists across all basic emotions except sadness. The effect size, however, is small, and previous studies have been underpowered.

Keywords: Major Depressive Disorder; Emotion Recognition; Meta-Analysis.

Meta-Analysis of Emotion Recognition Deficits in Major Depressive Disorder.

Introduction

The perception of emotion from non-verbal cues is crucial to human social interaction. Many psychological disorders are characterised by deficits or biases in facial emotion recognition, including schizophrenia (Addington *et al.*, 2006), alcoholism (Philippot *et al.*, 1999), autism (Celani *et al.*, 1999), anxiety (Button *et al.*, 2013a), bipolar disorder (Derntl *et al.*, 2009), and depression (Rubinow and Post, 1992).

Affective disorders affect 21 million people in Europe alone and account for nearly half of the costs of all mental disorders (Andlin-Sobocki *et al.*, 2005). Understanding the role of emotion recognition is especially relevant to depression, as the impaired recognition of emotion has been associated with decreased satisfaction, support, and well-being of interpersonal relationships (Carton *et al.*, 1999). Critically, poor interpersonal relationships have been proposed as an important factor in both the etiology and maintenance of depression (Finch and Zautra, 1992, Platt *et al.*, 2013), and impaired emotion recognition may contribute to the interpersonal difficulties and avoidance seen in depression (Persad and Polivy, 1993). Since deficits in emotion recognition may contribute to the maintenance of depressive symptoms, investigating this relationship has important implications for existing cognitive behavioural interventions and the development of novel interventions.

Many studies have attempted to investigate the relationship between emotion recognition and depression over the last 30 years (see Bourke *et al.*, 2010 for a review). However, these have used various paradigms and stimulus sets, thus making the comparison of results across studies difficult. Two recent meta-analyses were conducted to investigate the association between MDD and emotion recognition. Demenescu *et al.* (2010) examined 8 studies and found that emotion recognition in depressed adults was moderately impaired

compared to controls. Given the small number of included studies, analyses stratified by emotion and analyses of study-level design characteristics were not conducted. Similarly, Kohler *et al.* (2011) identified a moderate deficit in emotion recognition in a meta-analysis of 51 studies of emotion identification or discrimination in bipolar (31 studies) or unipolar (20 studies) depression patients compared to controls. Notably, impairment did not differ between diagnostic groups, and analyses of all six basic emotions revealed small to moderate deficits across both patient groups. However, data on specific emotions were limited, so it was difficult to determine with certainty whether the nature and strength of the deficit differed by emotion. There was also some evidence suggesting that symptom severity was associated with a greater deficit in emotion recognition. Furthermore, demographic characteristics such as old age, sex (females), and higher levels of education (in cases) were also shown to be positively associated with emotion recognition performance.

The results from these meta-analyses are inconclusive regarding whether the emotion recognition deficit in depression is general or specific to the recognition of one or more emotions. The discovery of a specific emotion recognition deficit in depression would have important implications for treatment, allowing clinicians to target the treatment of impairments more effectively. Some researchers have proposed that there is a unique relationship between major depressive disorder and the recognition of happiness, suggesting that the recognition of happiness is specifically impaired while the recognition of sadness is spared or enhanced (Bourke *et al.*, 2010, Gur *et al.*, 1992). Similarly, while studies have demonstrated that some antidepressant pharmacotherapies modify the recognition of emotion (Harmer *et al.*, 2011, Harmer *et al.*, 2013), meta-analyses thus far have not considered the effects of current medication on emotion recognition in depressed individuals.

The purpose of this meta-analysis was therefore to extend our understanding of the relationship between emotion recognition deficits and major depressive disorder. We did this

by comparing studies across several different methodologies, paradigms, and design-level characteristics, including stimulus sets, presentation times, and response options. This included stratifying our analyses by medication status (i.e., medicated or unmedicated) in order to investigate the effects of antidepressants on this relationship. In addition to investigating a general deficit of emotion recognition, we further stratified our analyses by all six basic emotions in order to investigate specific deficits. In the interest of reducing the moderate levels of heterogeneity detected in the previous meta-analyses, we only included studies using human facial emotional expression stimuli. Finally, we also calculated the statistical power of each study included in our analysis to detect the effect size indicated by the meta-analysis, and tested for possible publication bias. This meta-analysis extends our understanding of the relationship between emotion recognition abilities, and provides a more accurate estimate of the real magnitude of the effect of depression on emotion recognition deficits in studies using photorealistic stimuli.

Methods

Study Inclusion / Exclusion Criteria

Eligibility criteria for study inclusion were as follows: (1) studies were required to have both a clinical sample with a diagnosis of major depressive disorder (MDD) and a control sample; (2) studies were required to have assessed the accuracy of emotion recognition; and (3) studies were required to have used stimuli comprising of human facial emotional expressions. Studies using schematic or artistically rendered faces, neuroimaging studies and studies that included experimental administration of drug treatments were excluded. Studies that recruited participants with a diagnosis of both MDD and bipolar disorder were retained.

Search Strategy

We performed a search on two databases: PubMed and Web of Science. These databases were searched from the first date available in each database up to 1st June 2013, using the inclusion terms "depression", "MDD", "emotion*", "recognition", "perception" and the exclusion term "administration". After articles had been collected, bibliographies were then searched for additional references.

Data Extraction

For each study, the following data were extracted: (1) author (s) and year of publication; (2) data (mean and standard deviation of emotion recognition accuracy scores, number of participants, mean age and male/female ratio) and (3) study design characteristics. Study design was coded (where possible/applicable) for: stimulus emotion (anger, disgust, fear, happiness, sadness, surprise), case status (MDD no comorbidity, MDD comorbidity, MDD + bipolar disorder), control status (matched, unmatched), treatment status (medicated, unmedicated), diagnostic criteria (Diagnostic and Statistical Manual of Mental Disorders (DSM)/Research Diagnostic Criteria (RDC), International Classification of Diseases (ICD)), stimuli (Ekman & Friesen, Other), use of morphed stimuli (no, yes), stimulus type (dynamic, static), presentation time (<500 ms, >500 ms, 500 ms, self-paced), and response option (2 alternative forced choice (AFC), 6 AFC, other). We also rated the quality of all included studies using 8 items adapted from the Newcastle-Ottawa Scale, a measure for assessing the quality of nonrandomised studies in meta-analyses (Wells *et al.*, 2000). Studies were rated on the selection of study groups, the comparability of those groups, and the ascertainment of the outcome of interest.

Data Analysis

Effect sizes (Hedges's *g*) were calculated for the comparison of cases vs. controls on emotion recognition accuracy for each emotion reported within each individual study. Hedges's *g* is a measure of standardized mean difference, similar to Cohen's *d* but including a correction for small sample size. Conventionally, a small effect size is defined as .20, a medium effect size as .50 and a large effect size as .80 (Cohen, 1988).

Data were analyzed within a random effects framework, with *g*s pooled using DerSimonian and Lair (1986) methods. A random effects framework assumes that between-study variation is due to both chance or random variation and an individual study effect, and provides an estimate of the range of likely effect sizes across the populations sampled by individual studies. Random-effects models are more conservative than fixed-effects models and generate a wider confidence interval (CI), but give similar results under conditions of low between-study heterogeneity. The significance of the pooled *g* was determined using a *Z*-test. Between-study heterogeneity was estimated using the I^2 statistic. Conventionally, values of 25%, 50% and 75% represent the upper thresholds for low, moderate and high heterogeneity, respectively.

Small study bias, which may reflect publication bias against null results, was assessed using Egger's test (Egger *et al.*, 1997). We also conducted a series of stratified analyses and meta-regression analyses to assess the impact of various study design characteristics. The analyses were conducted using the Comprehensive Meta-Analysis (v2) statistical software package (Biostat, Englewood, NJ, USA). Exact *p*-values are reported throughout.

Results

Description of Studies

Our search strategy across both databases initially identified 728 articles. Of these, 66 articles were identified as duplicates and were removed. Of the remaining 662 articles, we

were able to exclude 624 articles because they did not meet our inclusion criteria. A further 16 articles were excluded because they did not report the data required to enable inclusion in our meta-analysis, and attempts to contact the study authors to acquire these were unsuccessful.

A total of 22 studies published between 1992 and 2012 met inclusion criteria and were included in our meta-analysis. Studies included in the analysis are marked in the bibliography with an asterisk (*). A flow chart describing this process is shown in Figure 1. Characteristics of these studies are described in Table 1.

Insert Table 1 and Figure 1 about here.

Quality of Included Studies

The 8 items we used to assess the quality of our included studies consisted of 4 items related to study group selection, 2 items related to the comparability of groups and 2 items related to how the studies ascertained the outcome of interest. Each study scored one point for each item if the criterion was met. Most studies included in our meta-analysis adequately described the selection of study groups as only three studies scored less than 3 out of a possible 4 points on these items. Most studies failed to offer sufficient information regarding the comparability of study groups as only six studies scored points on both items while three studies earned only 1 point. All studies met criteria regarding the ascertainment of the outcome of interest, scoring a point for both items.

Emotion Recognition in Major Depressive Disorder

Meta-analysis (k = 22) indicated strong evidence of a deficit in emotion recognition among cases compared to controls (g = -0.16, 95% CI -0.25 to -0.07, $p < 0.001$) with

negligible between-study heterogeneity (I^2 = 0%). Stratified analyses by across the six primary emotions indicated a deficit in emotion recognition for anger, disgust, fear, happiness and surprise (*k*s = 7 to 22, *g*s = -0.42 to -0.17, *p*s < 0.08), but not sadness (*k* = 21, *g* = -0.09, 95% CI -0.23 to +0.06, *p* = 0.23). Sensitivity analysis indicated that no single study disproportionately contributed to these results. These results are presented in Table 2.

Insert Table 2 about here.

Impact of Study-Level Design Characteristics

Stratified analyses indicated no evidence that any study-level design characteristics altered the deficit in emotion recognition among cases compared to controls (*p*s ≥ 0.11). In all cases, between-study heterogeneity was moderate to negligible ($I^2 \leq 55\%$), with the exception of the 2 studies in the MDD + bipolar disorder stratum (I^2 = 70%). These results are presented in Table 3. Meta-regression indicated a positive association between year of publication and effect size estimate (*p* = 0.029).

Insert Table 3 about here.

Impact of Medication Status on Recognition of Happiness and Sadness

Given evidence from human psychopharmacology studies indicating that antidepressants modify emotion recognition (Harmer *et al.*, 2013, Harmer *et al.*, 2011), we examined the impact of medication status on the recognition of happiness and sadness. The pattern of results described did not differ by medication status for either the recognition of happiness (*p* = 0.84) or sadness (*p* = 0.65). Notably, only 3 studies included in our analysis tested unmedicated cases compared to 19 studies assessing recognition of happiness in

medicated samples and 18 assessing recognition of sadness. These results are presented in Table 4.

Insert Table 4 about here.

Small Study Bias

There was evidence of small study bias for the combined analysis ($p = 0.003$), while for the stratified analyses this was indicated for sadness ($p = 0.028$) and anger ($p = 0.003$). Adjusting for possible publication bias against null results using Duval and Tweedie's trim and fill method (Duval and Tweedie, 2000) indicated a reduced effect size estimate in the combined analysis (g = -0.08, 95% CI -0.18 to +0.01), and the sadness (g = +0.04, 95% CI -0.12 to +0.20) and anger (g = 0.01, 95% CI -0.18 to +0.16) stratified analyses.

Power Analysis

The effect size estimate indicated by our combined meta-analysis (g = -0.16) suggests that a sample size of approximately 615 cases and 615 controls would be required to detect a deficit in emotion recognition with 80% power at an alpha level of 0.05. The median sample size among studies included in our meta-analysis was 21 cases and 25 controls, which would correspond to 8% power to detect an effect size of this magnitude.

Discussion

Our findings indicate a general emotion recognition deficit associated with major depressive disorder. In addition, analyses stratified by emotion indicate that the recognition of sadness is uniquely preserved, while recognition of the other basic emotions is impaired. We also did not find any evidence that study-level characteristics modified these results,

suggesting that these effects may be relatively robust to diagnostic criteria, task parameters and other design factors. Given the variability in these factors across studies, this finding was unexpected, and may suggest that despite the small effect size of the emotion recognition deficit, this is a robust feature of major depressive disorder.

Medication status amongst cases did not appear to modify the association of depression with emotion recognition. However, this analysis included only three studies where depressed patients were unmedicated at time of testing, making it difficult to draw firm conclusions on the effects of medication on emotion recognition in this population. Details of current psychological treatment, which may also modify emotion recognition, were often unreported and therefore could not be systematically examined. It is noteworthy that studies of medicated patients tended to report a greater deficit in the recognition of sadness than the studies including only unmedicated patients though there was not sufficient statistical power to evaluate whether this was a consistent effect. Clearly further research with untreated depressed samples is required in order to better understand how emotion recognition deficits are associated with major depressive disorder, rather than medication or therapy *per se*. In particular, a wide body of research suggests that antidepressant medication reduces the recognition of, and neural responses to, negative facial expressions in healthy participants and patients with major depressive disorder (see Pringle *et al.*, 2013). In the relative absence of data from unmedicated patients, it is possible that the current results are a marker of medication status as oppose to the disorder itself. Nonetheless, the results of our analyses stratified by emotion suggests that there is no unique relationship between major depressive disorder and the accurate recognition of happiness, as previously proposed (Bourke *et al.*, 2010, Gur *et al.*, 1992). Instead, the impaired recognition of happiness is merely part of a general recognition deficit across the other basic emotions. Response bias (i.e. the tendency

to label ambiguous faces as positive vs negative) was not systematically reported in these studies and therefore has not been directly compared.

Contemporary theories of depression emphasise the importance of negative biases in emotion recognition as an important causal factor in illness aetiology (Disner *et al.*, 2011, Roiser *et al.*, 2012). In particular, attentional, perceptual and interpretative biases towards negative material is believed to fuel negative self-referent schema in depression (Roiser *et al*., 2012). The current results are broadly consistent with this framework, since the recognition of sadness was preserved across a general landscape of emotion recognition deficits in depression. In other words, the recognition of sadness may be greater in relative terms, compared to the other emotional inputs (including happiness). However, the current results are not consistent with a more general negativity bias in depression in terms of accuracy of facial expression recognition. Based on the findings of a recent study, a negative bias in the interpretation of neutral faces rather than accuracy deficits in emotion recognition may represent a vulnerability factor for major depression in at-risk individuals (Maniglio *et al.*, 2014). Given the effects of medication on the detection of negative emotion in facial expressions (Harmer *et al.*, 2004, Harmer *et al.*, 2006), this conclusion needs to be qualified by noting the scarcity of research investigating emotion recognition in unmedicated patients. Future research should prioritise assessing emotion recognition (and associated measures) in patients free of medication. It is also worth noting that psychological treatments may also impact the processing of emotion in facial expressions, which indicates that studies in patients who are receiving neither pharmacological nor psychological treatments may be informative.

While our results indicate that major depressive disorder is associated with a general deficit in emotion recognition, the size of this effect is small. One consequence of this is the low statistical power of individual studies in our meta-analysis to detect effects of these

associations, with the largest study achieving only 34% power to detect the effect size indicated by our meta-analysis. The problems associated with low statistical power have recently been described, and include an increased likelihood that a statistically significant finding reflects a false positive (Button *et al.*, 2013b, Ioannidis, 2005). Rather than being endemic to a particular domain, the problem of low statistical power appears to be pervasive across several fields in the biomedical sciences. Our results therefore indicate the need for studies of emotion recognition deficits on a scale far larger than has been achieved to date. New technologies and data collection methods, such as the use of Internet and smartphone platforms, could help achieve this (Mar *et al.*, 2013), and recent studies have shown that data collected via the Mechanical Turk are of comparable fidelity to those collected in a traditional laboratory setting (Crump *et al.*, 2013). However, one important limitation of this approach is that it may be difficult to obtain data on clinical status except via self-report.

Our positive test for small study bias reveals evidence of possible publication bias against null results in this literature. This arises when researchers decide to not submit negative findings for publication, largely due to the prevailing tendency for journals to reject papers reporting null findings (Thornton and Lee, 2000). While we sought unpublished studies, as is common practice in conducting meta-analyses, we did not receive any responses. Given the presence of small study bias, we adjusted using Duval and Tweedie's trim and fill method (Duval and Tweedie, 2000) in order to account for small-study effects, where smaller studies in a meta-analysis tend to show larger treatment effects (Sterne *et al.*, 2000). While reduced in strength, evidence of a general deficit in emotion recognition in depression remained.

There are a number of limitations to the present study which should be considered when interpreting these results. First, we excluded studies using stimulus sets generated using schematic or artistically-rendered faces. This was done due to the lack of perceived

ecological validity for schematic or artistically-rendered faces, compared to human facial expression stimuli. We therefore cannot say whether our results would apply to tasks using schematic or artistically-rendered faces. Second, there was insufficient data among studies included in our analysis to conduct meta-analysis on response bias. The investigation of false alarms from recognition tasks would offer further insight into the nature of the observed deficits, and would allow us to potentially identify biased responding for specific emotions i.e. the tendency to mislabel ambiguous faces as sad or happy. However, there were minimal false alarm data available for analysis in the present study; future studies should report false alarm data consistently, alongside accuracy data. Third, we were limited in our ability to draw conclusions on the effect of medication status on emotion recognition in depressed individuals as few studies in our analysis assessed unmedicated cases. While the data available did not provide strong evidence that the recognition of emotion differs by medication status, contrary to our expectations given the literature on the effects of antidepressants on emotion recognition, future studies explicitly designed to test this (i.e., including both medicated and unmedicated cases) are required. Additionally, we were unable to investigate whether symptom severity moderated recognition performance as these data were not uniformly reported. As elevated depressive symptoms appear to predict poorer performance on recognition tasks, (Kohler *et al.*, 2011) accounting for symptom severity would be helpful when investigating the effect of medication status on performance. Finally, while we have determined a more accurate estimation of the size of the effect of depression on emotion recognition performance, it is not clear how these effects may translate to clinical significance. Therefore more research is needed to explore the relationship between emotion recognition and symptom severity.

In conclusion, our analyses confirm a general deficit of emotion recognition in depressed individuals compared to controls, albeit with a small effect size. Studies thus far

have been considerably underpowered to detect this effect, and primary studies with much larger sample sizes will be required to properly investigate this association. Of the six basic emotions, only the recognition of sadness appears to be spared in depression, and there appears to be no specific association with impaired recognition of happiness. Further research comparing both medicated and unmedicated patients would offer new insights into the effects of depression on emotion recognition, with possible implications for treatment.

Acknowledgements

M.R.M. and M.N.D. are members of the UK Centre for Tobacco Control Studies, a UKCRC Public Health Research: Centre of Excellence. Funding from British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, and the National Institute for Health Research, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged.

Conflict of Interest

I.S.P-V. and M.R.M. are directors of Jericoe Ltd, which develops software for assessing and modifying emotion perception. M.R.M. has provided consultancy to Servier. CJH is a director of Oxford Psychologists and has received consultancy fees from Lundbeck, Lily, Servier and P1vital in the last three years. She holds shares in P1vital.

References

Addington, J., Saeedi, H. & Addington, D. (2006). Facial affect recognition: A mediator between cognitive and social functioning in psychosis? *Schizophrenia Research* **85**, 142-150.

***Anderson, I. M., Shippen, C., Juhasz, G., Chase, D., Thomas, E., Downey, D., Toth, Z. G., Lloyd-Williams, K., Elliott, R. & Deakin, J. F. W.** (2011). State-dependent alteration in face emotion recognition in depression. *British Journal of Psychiatry* **198**, 302-308.

Andlin-Sobocki, P., Jönsson, B., Wittchen, H.-U. & Olesen, J. (2005). Cost of disorders of the brain in Europe. *European Journal of Neurology* **12**, 1-27.

***Arteche, A., Joormann, J., Harvey, A., Craske, M., Gotlib, I. H., Lehtonen, A., Counsell, N. & Stein, A.** (2011). The effects of postnatal maternal depression and anxiety on the processing of infant faces. *Journal of Affective Disorders* **133**, 197-203.

***Bediou, B., Krolak-Salmon, P., Saoud, M., Henaff, M. A., Burt, M., Dalery, J. & D'Amato, T.** (2005). Facial expression and sex recognition in schizophrenia and depression. *Canadian Journal of Psychiatry-Revue Canadienne De Psychiatrie* **50**, 525-533.

Bourke, C., Douglas, K. & Porter, R. (2010). Processing of facial emotion expression in major depression: a review. *Australian & New Zealand Journal of Psychiatry* **44**, 681-96.

Button, K., Lewis, G., Penton-Voak, I. & Munafo, M. (2013a). Social anxiety is associated with general but not specific biases in emotion recognition. *Psychiatry Research* **210**, 199-207.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J. & Munafo, M. R. (2013b). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience* **14**, 365-376.

Carton, J., Kessler, E. & Pape, C. (1999). Nonverbal decoding skills and relationship well-being in adults. *Journal of Nonverbal Behavior* **23**, 91-100.

Catala-Lopez, F., Genova-Maleras, R., Vieta, E. & Tabares-Seisdedos, R. (2013). The increasing burden of mental and neurological disorders. *European Neuropsychopharmacology* **23**, 1337-9.

Celani, G., Battacchi, M. W. & Arcidiacono, L. (1999). The understanding of the emotional meaning of facial expressions in people with autism. *Journal of Autism and Developmental Disorders* **29**, 57-66.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. L. Erlbaum Associates: Hillsdale, NJ.

Crump, M. J. C., McDonnell, J. V. & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE* **8**, e57410.

Demenescu, L. R., Kortekaas, R., den Boer, J. A. & Aleman, A. (2010). Impaired Attribution of Emotion to Facial Expressions in Anxiety and Major Depression. *PLoS ONE* **5**, e15058.

***Derntl, B., Seidel, E.-M., Kryspin-Exner, I., Hasmann, A. & Dobmeier, M.** (2009). Facial emotion recognition in patients with bipolar I and bipolar II disorder. *British Journal of Clinical Psychology* **48**, 363-375.

Derntl, B., Seidel, E. M., Schneider, F. & Habel, U. (2012). How specific are emotional deficits? A comparison of empathic abilities in schizophrenia, bipolar and depressed patients. *Schizophrenia Research*.

DerSimonian, R. & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**, 177-188.

Disner, S. G., Beevers, C. G., Haigh, E. A. & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. *Nature Reviews Neuroscience* **12**, 467-77.

***Douglas, K. M. & Porter, R. J.** (2010). Recognition of disgusted facial expressions in severe depression. *British Journal of Psychiatry* **197**, 156-7.

Duval, S. & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* **56**, 455-63.

Egger, M., Davey Smith, G., Schneider, M. & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* **315**, 629-34.

Finch, J. F. & Zautra, A. J. (1992). Testing latent longitudinal models of social ties and depression among the elderly: a comparison of distribution-free and maximum likelihood estimates with nonnormal data. *Psycholollgy and Aging* **7**, 107-18.

***Gaebel, W. & Wolwer, W.** (1992). Facial expression and emotional face recognition in schizophrenia and depression. *European Archives of Psychiatry and Clinical Neuroscience* **242**, 46-52.

***Gollan, J. K., McCloskey, M., Hoxha, D. & Coccaro, E. F.** (2010). How do depressed and healthy adults interpret nuanced facial expressions? *Journal of Abnormal Psychology* **119**, 804-10.

***Gur, R. C., Erwin, R. J., Gur, R. E., Zwil, A. S., Heimberg, C. & Kraemer, H. C.** (1992). Facial emotion discrimination: 2. Behavioral findings in depression. *Psychiatry Research* **42**, 241-251.

Harmer, C. J., Dawson, G. R., Dourish, C. T., Favaron, E., Parsons, E., Fiore, M., Zucchetto, M., Bifone, A., Poggesi, I., Fernandes, S., Alexander, R. C. & Goodwin, G. M. (2013). Combined NK1 antagonism and serotonin reuptake inhibition: effects on emotional processing in humans. *Journal of Psychopharmacology* **27**, 435-443.

Harmer, C. J., de Bodinat, C., Dawson, G. R., Dourish, C. T., Waldenmaier, L., Adams, S., Cowen, P. J. & Goodwin, G. M. (2011). Agomelatine facilitates positive versus negative affective processing in healthy volunteer models. *Journal of Psychopharmacology* **25**, 1159-1167.

Harmer, C. J., Mackay, C. E., Reid, C. B., Cowen, P. J. & Goodwin, G. M. (2006). Antidepressant drug treatment modifies the neural processing of nonconscious threat cues. *Biological Psychiatry* **59**, 816-820.

Harmer, C. J., Shelley, N. C., Cowen, P. J. & Goodwin, G. M. (2004). Increased positive versus negative affective perception and memory in healthy volunteers following selective serotonin and norepinephrine reuptake inhibition. *American Journal of Psychiatry* **161**, 1256-1263.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Med* **2**, e124.

***Joormann, J. & Gotlib, I. H.** (2006). Is this happiness I see? Biases in the identification of emotional facial expressions in depression and social phobia. *Journal of Abnormal Psychology* **115**, 705-14.

***Kan, Y., Mimura, M., Kamijima, K. & Kawamura, M.** (2004). Recognition of emotion from moving facial and prosodic stimuli in depressed patients. *Journal of Neurology, Neurosurgery & Psychiatry* **75**, 1667-1671.

Kohler, C. G., Hoffman, L. J., Eastman, L. B., Healey, K. & Moberg, P. J. (2011). Facial emotion perception in depression and bipolar disorder: a quantitative review. *Psychiatry Research* **188**, 303-9.

***Langenecker, S. A., Caveney, A. F., Giordani, B., Young, E. A., Nielson, K. A., Rapport, L. J., Biellauskas, L. A., Mordhorst, M. J., Marcus, S., Yodkovik, N., Kerber, K., Berent, S. & Zubieta, J. K.** (2007). The sensitivity and psychometric properties of a brief computer-based cognitive screening battery in a depression clinic. *Psychiatry Research* **152**, 143-154.

***Leppanen, J. M., Milders, M., Bell, J. S., Terriere, E. & Hietanen, J. K.** (2004). Depression biases the recognition of emotionally neutral faces. *Psychiatry Research* **128**, 123-33.

***Mah, L. & Pollock, B. G.** (2010). Emotional processing deficits in late-life depression. *The American Journal of Geriatric Psychiatry* **18**, 652-6.

Maniglio, R., Gusciglio, F., Lofrese, V., Belvederi Murri, M., Tamburello, A. & Innamorati, M. (2014). Biased processing of neutral facial expressions is associated with depressive symptoms and suicide ideation in individuals at risk for major depression due to affective temperaments. *Comprehensive Psychiatry* **55**, 518-25.

Mar, R. A., Spreng, R. N. & Deyoung, C. G. (2013). How to produce personality neuroscience research with high statistical power and low additional cost. *Cognitive, Affective & Behavioral Neuroscience*.

***Mendlewicz, L., Linkowski, P., Bazelmans, C. & Philippot, P.** (2005). Decoding emotional facial expressions in depressed and anorexic patients. *Journal of Affective Disorders* **89**, 195-9.

***Milders, M., Bell, S., Platt, J., Serrano, R. & Runcie, O.** (2010). Stable expression recognition abnormalities in unipolar depression. *Psychiatry Research* **179**, 38-42.

Miret, M., Ayuso-Mateos, J. L., Sanchez-Moreno, J. & Vieta, E. (2013). Depressive disorders and suicide: Epidemiology, risk factors, and burden. *Neuroscience & Biobehavioral Reviews*.

***Naranjo, C., Kornreich, C., Campanella, S., Noel, X., Vandriette, Y., Gillain, B., de Longueville, X., Delatte, B., Verbanck, P. & Constant, E.** (2011). Major depression is associated with impaired processing of emotion in music as well as in facial and vocal stimuli. *Journal of Affective Disorders* **128**, 243-251.

***Persad, S. M. & Polivy, J.** (1993). Differences between depressed and nondepressed individuals in the recognition of and response to facial emotional cues. *Journal of Abnormal Psychology* **102**, 358-68.

Philippot, P., Kornreich, C., Blairy, S., Baert, I., Dulk, A. D., Bon, O. L., Streel, E., Hess, U., Pelc, I. & Verbanck, P. (1999). Alcoholics' deficits in the decoding of emotional facial expression. *Alcoholism: Clinical and Experimental Research* **23**, 1031-1038.

Platt, B., Kadosh, K. C. & Lau, J. Y. F. (2013). The role of peer rejection in adolescent depression. *Depression and Anxiety* **30**, 809-821.

Pringle, A., McCabe, C., Cowen, P. J. & Harmer, C. J. (2013). Antidepressant treatment and emotional processing: can we dissociate the roles of serotonin and noradrenaline? *Journal of Psychopharmacology* **27**, 719-31.

Roiser, J. P., Elliott, R. & Sahakian, B. J. (2012). Cognitive mechanisms of treatment in depression. *Neuropsychopharmacology* **37**, 117-136.

Rubinow, D. R. & Post, R. M. (1992). Impaired recognition of affect in facial expression in depressed patients. *Biological Psychiatry* **31**, 947-953.

Salovey, P. & Grewal, D. (2005). The science of emotional intelligence. *Current Directions in Psychological Science* **14**, 281-285.

***Schaefer, K. L., Baumann, J., Rich, B. A., Luckenbaugh, D. A. & Zarate, C. A., Jr.** (2010). Perception of facial emotion in adults with bipolar or unipolar depression and controls. *Journal of Psychiatric Research* **44**, 1229-35.

***Schepman, K., Taylor, E., Collishaw, S. & Fombonne, E.** (2012). Face emotion processing in depressed children and adolescents with and without comorbid conduct disorder. *Journal of Abnormal Child Psychology* **40**, 583-93.

***Sprengelmeyer, R., Steele, J. D., Mwangi, B., Kumar, P., Christmas, D., Milders, M. & Matthews, K.** (2011). The insular cortex and the neuroanatomy of major depression. *Journal of Affective Disorders* **133**, 120-127.

Sterne, J. A. C., Gavaghan, D. & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology* **53**, 1119-1129.

Thornton, A. & Lee, P. (2000). Publication bias in meta-analysis: its causes and consequences. *Journal of Clinical Epidemiology* **53**, 207-216.

***Vederman, A. C., Weisenbach, S. L., Rapport, L. J., Leon, H. M., Haase, B. D., Franti, L. M., Schallmo, M. P., Saunders, E. F., Kamali, M. M., Zubieta, J. K., Langenecker, S. A. & McInnis, M. G.** (2012). Modality-specific alterations in the perception of emotional stimuli in bipolar disorder compared to healthy controls and major depressive disorder. *Cortex* **48**, 1027-34.

Wells, G.A., Shea, B., O'Connell, D., Peterson, J., Welch, V., Tugwell, P. (2000, July). The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses. Paper presented at the 3rd symposium on systematic reviews: beyond the basics, Oxford.

***Wright, S. L., Langenecker, S. A., Deldin, P. J., Rapport, L. J., Nielson, K. A., Kade, A. M., Own, L. S., Akil, H., Young, E. A. & Zubieta, J. K.** (2009). Gender-specific disruptions in emotion processing in younger adults with depression. *Depress Anxiety* **26**, 182-9.

Figure 1. Flow diagram of search results.

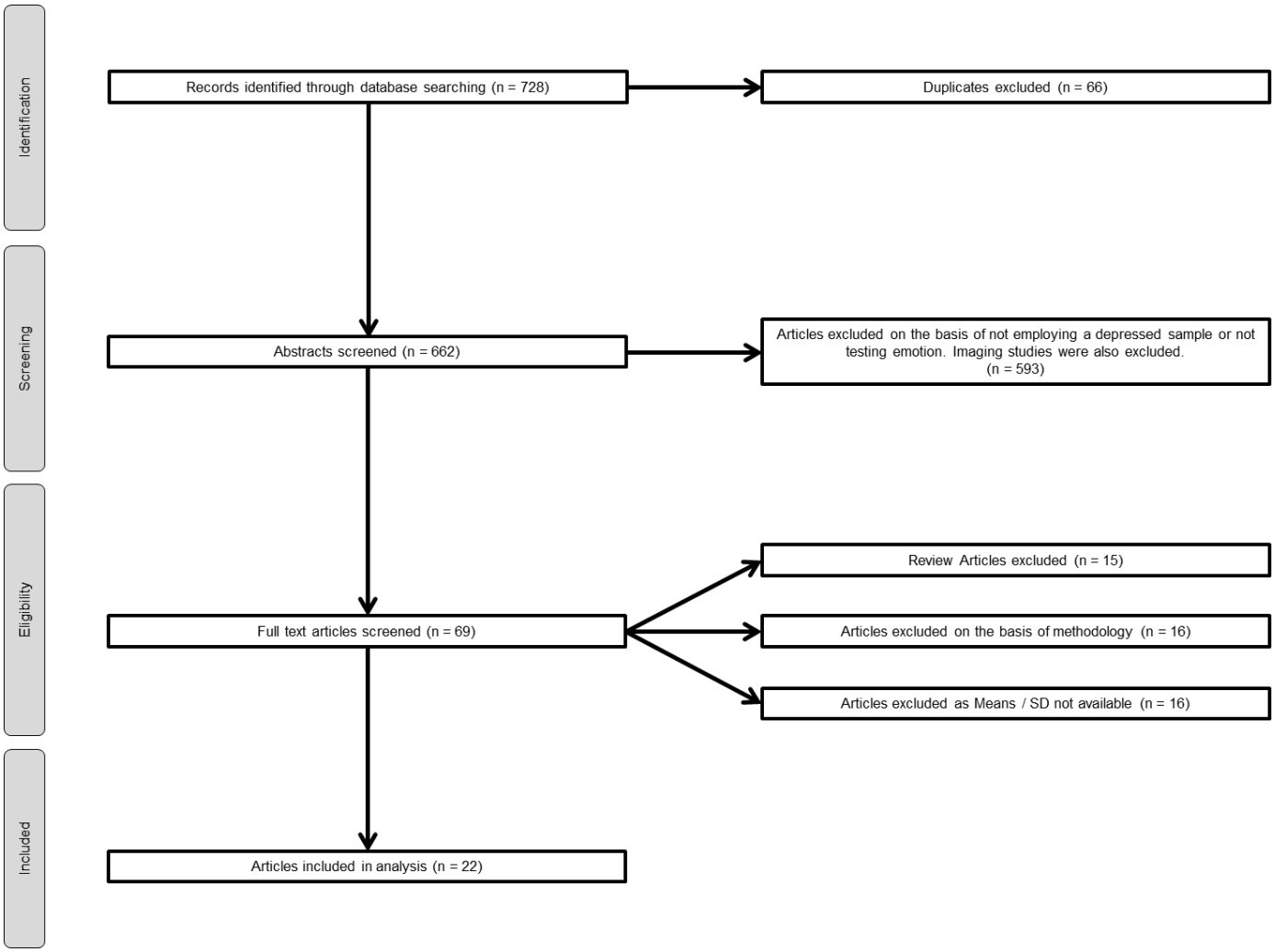

Table 1. Characteristics of Included Studies.

Author	Year	Anger	Disgust	Fear	Happiness	Sadness	Surprise	N (cases)	N (controls)	Effect (g)	Power	Treatment	Cases	Controls	Diagnosis	Stimuli	Stimuli Morphed	Stimulus Type	Presentation Time	Response Option	Age (cases)	Age (controls)	% Female (cases)	% Female (controls)
Anderson	2011	x	x	x	x	x	x	30	101	-0.02	13%	Medicated	Comorbidity	Unmatched	DSM/RDC	Ekman/Friesen	Yes	Static	>500 ms	Other	39	30	73	57
Arteche	2011				x	x		21	34	-0.40	9%	Medicated	Comorbidity	Unmatched	DSM/RDC	Other	Yes	Dynamic	500 ms	2 AFC	32	34	100	100
Bediou	2005		x	x	x			20	20	-0.12	8%	Medicated	No Comorbidity	Unmatched	DSM/RDC	Other	Yes	Static	<500 ms	Other	39	26	35	35
Derntl	2012	x	x	x	x	x		24	24	-0.21	9%	Medicated	No Comorbidity	Matched	DSM/RDC	Other	No	Static	>500 ms	2 AFC	41	40	50	50
Douglas	2010	x	x	x	x	x		68	50	-0.30	15%	Medicated	Bipolar Disorder	Unmatched	DSM/RDC	Ekman/Friesen	Yes	Static	500 ms	6 AFC	40	39	59	63
Gaebel	1992	x	x	x	x	x	x	21	15	-0.02	8%	Medicated	No Comorbidity	Unmatched	DSM/RDC	Ekman/Friesen	No	Static	>500ms	Other	39	31	43	40
Gollan	2010	x	x	x	x	x	x	44	44	-0.09	12%	Drug-Free	No Comorbidity	Unmatched	DSM/RDC	Ekman/Friesen	Yes	Static	500 ms	6 AFC	28	31	57	68
Gur	1992				x	x		14	14	-1.11	7%	Medicated	Bipolar Disorder	Matched	DSM/RDC	Other	No	Static	>500 ms	Other	45	37	86	86
Joormann	2006	x		x	x	x		21	25	-0.16	9%	Medicated	Comorbidity	Unmatched	DSM/RDC	Ekman/Friesen	Yes	Dynamic	500 ms	Other	34	32	86	68
Kan	2004	x	x	x	x	x	x	16	20	-0.10	8%	Medicated	No Comorbidity	Unmatched	DSM/RDC	Other	No	Dynamic	>500 ms	6 AFC	51	59	44	50
Langenecker	2007	x		x	x	x		200	71	-0.14	23%	Medicated	Comorbidity	Unmatched	DSM/RDC	Ekman/Friesen	No	Static	<500 ms	Other	35	25	68	57
Leppanen	2004				x	x		18	18	-0.22	8%	Medicated	No Comorbidity	Matched	ICD	Ekman/Friesen	No	Static	<500 ms	Other	45	45	61	61
Mah	2010			x	x	x		11	11	-0.11	7%	Drug-Free	No Comorbidity	Unmatched	DSM/RDC	Other	No	Static	Self-paced	Other	73	75	64	73
Mendlewicz	2005	x	x	x	x	x		21	32	-0.62	9%	Medicated	No Comorbidity	Unmatched	DSM/RDC	Other	Yes	Static	Self-paced	Other	17	21	100	100
Milders	2010	x	x	x	x	x		19	25	-0.06	8%	Medicated	Comorbidity	Matched	ICD	Ekman/Friesen	Yes	Static	500 ms	6 AFC	46	48	58	72
Naranjo	2011	x		x	x	x		23	23	-0.54	9%	Medicated	No Comorbidity	Matched	DSM/RDC	Other	No	Static	Self-paced	Other	41	40	78	78
Persad	1993	x	x	x	x	x	x	16	16	-0.51	8%	Medicated	No Comorbidity	Unmatched	DSM/RDC	Ekman/Friesen	No	Static	Self-paced	Other	n/a	n/a	100	100
Schaefer	2010	x	x	x	x	x	x	34	24	-0.21	10%	Drug-Free	No Comorbidity	Unmatched	DSM/RDC	Ekman/Friesen	Yes	Dynamic	<500 ms	6 AFC	45	45	44	50
Schepman	2012	x		x	x	x		29	37	-0.04	10%	Medicated	Comorbidity	Matched	DSM/RDC	Other	Yes	Static	>500 ms	Other	16	15	66	62
Sprengelmeyer	2011	x	x	x	x	x	x	10	45	-0.40	8%	Medicated	Comorbidity	Unmatched	DSM/RDC	Ekman/Friesen	Both	Static	>500 ms	6 AFC	50	51	70	70
Vederman	2012	x		x	x	x		78	66	-0.01	17%	Medicated	No Comorbidity	Unmatched	DSM/RDC	Ekman/Friesen	No	Static	<500 ms	Other	39	37	69	64
Wright	2009	x		x	x	x		239	128	-0.17	34%	Medicated	Comorbidity	Unmatched	DSM/RDC	Ekman/Friesen	No	Static	<500 ms	Other	26	25	71	47

Note. MDD = major depressive disorder; DSM = Diagnostic and Statistical Manual of Mental Disorders; RDC = Research Diagnostic Criteria; ICD = International Classification of Diseases; AFC = alternative forced choice.

Table 2. Meta-Analysis of Emotion Recognition in Major Depressive Disorder by Emotion.

	k	***g***	**Lower**	**Upper**	***P***	I^2	P_{Egger}
All studies	22	-0.162	-0.250	-0.074	<0.001	0%	0.003
Emotion							
Anger	16	-0.220	-0.376	-0.062	0.006	57%	0.003
Disgust	11	-0.420	-0.646	-0.195	<0.001	53%	0.69
Fear	17	-0.248	-0.372	-0.123	<0.001	35%	0.50
Happiness	22	-0.167	-0.255	-0.080	<0.001	0%	0.38
Sadness	21	-0.088	-0.234	+0.057	0.23	56%	0.028
Surprise	7	-0.170	-0.358	+0.018	0.076	0%	0.29

Table 3. Meta-Analysis of Emotion Recognition in Major Depressive Disorder by Study Design Characteristics.

	k	***g***	**Lower**	**Upper**	***P***	I^2	P_{diff}
Cases †							
MDD no comorbidity	12	-0.210	-0.363	-0.057	0.007	0%	0.27
MDD comorbidity	8	-0.104	-0.218	+0.009	0.071	0%	
MDD + bipolar disorder	2	-0.638	-1.452	+0.176	0.12	70%	
Controls							
Matched	6	-0.247	-0.546	+0.052	0.11	55%	0.71
Unmatched	16	-0.187	-0.291	-0.084	<0.001	0%	
Medication ‡							
Medicated	19	-0.170	-0.265	-0.075	<0.001	4%	0.83
Unmedicated	3	-0.134	-0.440	+0.172	0.39	0%	
Diagnostic Criteria							
DSM/RDC	20	-0.165	-0.256	-0.074	<0.001	0%	0.81
ICD	2	-0.119	-0.476	+0.238	0.51	0%	
Stimuli							
Ekman & Friesen	13	-0.163	-0.269	-0.056	0.003	0%	0.34
Other	9	-0.295	-0.544	-0.045	0.021	46%	
Stimuli Morphed §							
No	12	-0.197	-0.320	-0.075	0.002	0%	0.43
Yes	11	-0.126	-0.253	+0.001	0.051	0%	
Stimulus Type							
Dynamic	4	-0.235	-0.521	+0.052	0.11	0%	0.64
Static	18	-0.162	-0.259	-0.064	0.001	6%	
Presentation Time							
Self-paced	4	-0.504	-0.830	-0.177	0.003	0%	0.26
>500 ms	7	-0.182	-0.413	+0.050	0.13	42%	
500ms	5	-0.196	-0.397	+0.005	0.056	0%	
<500 ms	6	-0.140	-0.277	-0.002	0.046	0%	
Response Option							
2 AFC	2	-0.313	-0.708	+0.082	0.12	0%	0.11
4 AFC	7	-0.074	-0.191	+0.044	0.22	0%	
6 AFC	6	-0.201	-0.388	-0.013	0.036	0%	
Other	7	-0.162	-0.568	-0.144	0.001	26%	

Note. MDD = major depressive disorder; DSM = Diagnostic and Statistical Manual of Mental Disorders; RDC = Research Diagnostic Criteria; ICD = International Classification of Diseases; AFC = alternative forced choice.

† One study with cases classified as MDD+BP included participants diagnosed with comorbid disorders (Douglas and Porter, 2010).

‡ Studies classified as "medicated" include those where only a proportion of participants were medicated (Anderson *et al.*, 2011, Arteche *et al.*, 2011, Bediou *et al.*, 2005, Derntl *et al.*, 2012, Douglas and Porter, 2010, Gur *et al.*, 1992, Joormann and Gotlib, 2006, Kan *et al.*, 2004, Langenecker *et al.*, 2007, Mendlewicz *et al.*, 2005, Milders *et al.*, 2010, Vederman *et al.*, 2012, Wright *et al.*, 2009).

§ One study used both morphed and unmorphed stimuli in two separate tasks, and contributed to each stratum of this analysis (Sprengelmeyer *et al.*, 2011).

Table 4. Meta-Analysis of Happiness and Sadness Recognition in Major Depressive Disorder by Medication Status.

	k	***g***	**Lower**	**Upper**	***P***	I^2	P_{diff}
Happiness							
Medicated †	19	-0.164	-0.256	-0.073	<0.001	0%	0.84
Unmedicated	3	-0.197	-0.502	+0.108	0.21	0%	
Sadness							
Medicated †	18	-0.106	-0.260	+0.048	0.18	56%	0.65
Unmedicated	3	+0.026	-0.515	+0.568	0.92	63%	

† Studies classified as "medicated" include those where only a proportion of participants were medicated (Anderson *et al.*, 2011, Arteche *et al.*, 2011, Bediou *et al.*, 2005, Derntl *et al.*, 2012, Douglas and Porter, 2010, Gur *et al.*, 1992, Joormann and Gotlib, 2006, Kan *et al.*, 2004, Langenecker *et al.*, 2007, Mendlewicz *et al.*, 2005, Milders *et al.*, 2010, Vederman *et al.*, 2012, Wright *et al.*, 2009).

Supplementary Material

Quality of Studies

Table S1. *Assessment of quality of studies.*

Study (year)	Selection				Comparability		Outcome		Total
	1	2	3	4	5	6	7	8	
Anderson (2011)	☆	☆	☆	☆	☆	☆	☆	☆	8
Arteche (2011)	☆	☆	☆	☆	..	..	☆	☆	6
Bediou (2005)	☆	..	☆	☆	..	..	☆	☆	5
Derntl (2012)	☆	☆	☆	☆	☆	☆	☆	☆	8
Douglas (2010)	☆	☆	☆	☆	..	..	☆	☆	6
Gaebel (1992)	☆	☆	☆	..	..	..	☆	☆	5
Gollan (2010)	☆	☆	☆	☆	..	..	☆	☆	6
Gur (1992)	☆	☆	☆	☆	☆	..	☆	☆	7
Joormann (2006)	☆	☆	☆	☆	..	..	☆	☆	6
Kan (2004)	☆	..	..	☆	..	..	☆	☆	4
Langenecker (2007)	☆	☆	☆	☆	☆	☆	☆	☆	8
Leppanen (2004)	☆	☆	☆	☆	☆	☆	☆	☆	8
Mah (2010)	☆	☆	☆	☆	..	..	☆	☆	6
Mendlewicz (2005)	☆	☆	☆	☆	☆	..	☆	☆	7
Milders (2010)	☆	..	☆	☆	☆	☆	☆	☆	7
Naranjo (2011)	☆	☆	☆	☆	☆	☆	☆	☆	8
Persad (1993)	☆	☆	☆	☆		☆	☆	☆	7
Schaefer (2010)	☆	☆	☆	☆	..	..	☆	☆	6
Schepman (2012)	☆	☆	☆	☆	☆	☆	☆	☆	8

Sprengelmeyer (2011)	☆	☆	..	☆	..	..	☆	☆	5
Vederman (2012)	☆	..	..	..	..	..	☆	☆	3
Wright (2009)	☆	..	..	☆	..	..	☆	☆	4

Selection items 1 and 2 assessed the adequacy and representativeness of the cases respectively, while items 3 and 4 assessed the selection and definition of controls, respectively. Comparability items assessed whether the studies matched cases and controls for gender (item 5) and age (item 6). Outcome items 7 and 8 assessed the method used to ascertain outcomes of interest and whether the same method was used for both cases and controls, respectively. A star denotes that the study received a point for that item.

While Vederman et al. (2012) rated poorly on the scale, we found no evidence for a difference in our results when excluding this study from our analyses. Therefore we retained this study in our analyses.