
Peer reviewed version

License (if available):
CC BY-NC-ND

Link to published version (if available):
10.1016/j.schres.2019.12.031

Link to publication record in Explore Bristol Research

PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Elsevier at https://www.sciencedirect.com/science/article/pii/S0920996419306012?via%3Dihub. Please refer to any applicable terms of use of the publisher.
Title: Systematic review and meta-analysis of the relationship between genetic risk for schizophrenia and facial emotion recognition

Article Type: Full Length Article

Keywords: Schizophrenia; facial emotion recognition; genetics; family history; high risk.

Corresponding Author: Dr. David Martin, MBBS

Corresponding Author's Institution: University of Bristol

First Author: David Martin, MBBS

Order of Authors: David Martin, MBBS; Jazz Croft, MSc; Alice Pitt, MBBS; Daniela Strelchuk, MSc; Sarah Sullivan, PhD; Stan Zammit, PhD

Abstract: Background

Recent research has highlighted that facial emotion recognition deficits are more common in people with schizophrenia, but the reason for this association is not well understood. Comparing facial recognition deficits in unaffected individuals at higher genetic risk for schizophrenia with individuals at lower genetic risk could increase our understanding of this relationship.

Methods

We systematically reviewed studies reporting on the relationship between genetic risk of schizophrenia and facial emotion recognition deficits. Meta-analyses were performed where sufficient data were available, otherwise we conducted narrative summaries. Meta-analyses were performed both for generalised and specific facial emotion recognition deficits.

Results

34 studies were included in this review with 23 included in meta-analyses. Meta-analysis indicated strong evidence of a deficit in facial emotion recognition in first-degree relatives of people with schizophrenia compared with controls (SMD 0.38 95%CI 0.26 to 0.51, p < 0.001). Further meta-analyses demonstrated strong evidence of a deficit in the recognition of negative valence facial expressions (SMD 0.19 CI 0.06 to 0.32, p = 0.004) but no evidence of deficit in the recognition of neutral or positive valence.

Conclusions

There is strong evidence of facial emotion recognition deficits in first-degree relatives of people with schizophrenia. Our findings suggest that such deficits in people with schizophrenia arise prior to the onset of the disorder, though cannot inform whether that association is causal or
due to confounding. Emotion recognition deficits, particularly to negative emotions, might be useful predictors of schizophrenia risk.
Dear Editor,

Thank you for accepting our submission with minor changes. These have now been addressed and our manuscript re-submitted.

With kind regards,

Dr David Martin

Also on behalf of:
Professor Stan Zammit, Dr Sarah Sullivan, Ms Jazz Croft, Dr Alice Pitt and Ms Daniela Strelchuk.
Systematic review and meta-analysis of the relationship between genetic risk for schizophrenia and facial emotion recognition

Authors and institutional affiliations
Dr David Martin
Ms Jazz Croft
Dr Alice Pitt
Ms Daniela Strelchuk
Dr Sarah Sullivan
Professor Stan Zammit

a Centre for Academic Mental Health, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, UK
b Avon and Wiltshire Mental Health Partnership NHS Trust, UK
c MRC Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Hadyn Ellis Building, Maindy Road, Cathays, Cardiff, CF24 4HQ, UK

Corresponding author
Dr David Martin
Centre for Academic Mental Health, Population Health Sciences, Bristol Medical School, Oakfield House, Oakfield Grove, Bristol, BS8 2BN
Email: dave.martin@bristol.ac.uk

Word count
Abstract = 240
Text body = 3750
Abstract

Background
Recent research has highlighted that facial emotion recognition deficits are more common in people with schizophrenia, but the reason for this association is not well understood. Comparing facial recognition deficits in unaffected individuals at higher genetic risk for schizophrenia with individuals at lower genetic risk could increase our understanding of this relationship.

Methods
We systematically reviewed studies reporting on the relationship between genetic risk of schizophrenia and facial emotion recognition deficits. Meta-analyses were performed where sufficient data were available, otherwise we conducted narrative summaries. Meta-analyses were performed both for generalised and specific facial emotion recognition deficits.

Results
34 studies were included in this review with 23 included in meta-analyses. Meta-analysis indicated strong evidence of a deficit in facial emotion recognition in first-degree relatives of people with schizophrenia compared with controls (SMD 0.38 95%CI 0.26 to 0.51, p = <0.001). Further meta-analyses demonstrated strong evidence of a deficit in the recognition of negative valence facial expressions (SMD 0.19 CI 0.06 to 0.32, p = 0.004) but no evidence of deficit in the recognition of neutral or positive valance.

Conclusions
There is strong evidence of facial emotion recognition deficits in first-degree relatives of people with schizophrenia. Our findings suggest that such deficits in people with schizophrenia arise prior to the onset of the disorder, though cannot inform whether that association is causal or due to confounding. Emotion recognition deficits, particularly to negative emotions, might be useful predictors of schizophrenia risk.

Keywords
Schizophrenia, facial emotion recognition, genetic, family history, high risk.
1 Introduction

Schizophrenia is a psychotic disorder characterised by hallucinations, delusions, disorganized speech or behaviour, and impaired cognitive ability. It has a lifetime risk of about 1% (Schizophrenia Working Group of the Psychiatric Genomics, 2014), and carries a significant health, social and financial burden for the individual, people close to them, and the wider society (Knapp et al., 2004). Our limited understanding of the aetiology of schizophrenia means that therapeutic options are limited. Pharmacological interventions are the first line treatment, although psychological treatments aimed at addressing a person’s beliefs about their symptoms and at increasing adherence to medication are also used (Patel et al., 2014).

Recent research has highlighted that facial emotion recognition deficits are more prevalent in people with schizophrenia (Aleman and Kahn, 2005; Mandal, 1998) compared to people without this disorder. These are deficits in the recognition of the emotional state of another person by observation of their facial expression. Facial emotion recognition deficits in schizophrenia are associated with an extensive pattern of activation abnormalities on fMRI, consistent with hypoactive emotion recognition networks (Jani and Kasparek, 2018). Compensatory over-activation in the medial prefrontal cortex (MPFC) during threatening faces processing has also been demonstrated (Dong et al., 2017).

These deficits are present during the prodromal phase of the illness (Green et al., 2012), in people with first episode psychosis (Bosnjak Kuharic et al., 2019; Daros et al., 2014) and those with schizophrenia (Kohler et al., 2010; Savla et al., 2013). There is evidence to suggest that patients with schizophrenia have specific facial emotion recognition deficits in the recognition of negative emotions, particularly fear and anger, compared with neutral or positive emotions (Addington et al., 2006).

The reason for the association between facial emotion recognition deficits and schizophrenia is not well understood. It is possible that facial emotion recognition deficits occur secondary to schizophrenia or the association could be due to confounding, whereby
schizophrenia and facial emotion recognition deficits share genetic or environmental risk factors.

Schizophrenia has a heritability of around 80% (Cardno et al., 1999), and a family history of schizophrenia is one of the strongest risk factors for this disorder. Examining whether individuals who are at higher genetic risk for schizophrenia, but are unaffected, have an increased likelihood of facial emotion recognition deficits compared with individuals at lower genetic risk could increase our understanding of the relationship between these deficits and schizophrenia. Consistent evidence of association in such studies reduces the likelihood of reverse causation as an explanation, though they cannot discriminate between genetic confounding and causality.

A systematic review and meta-analysis conducted in 2012 (Lavoie et al., 2013) reported that first-degree relatives of people with schizophrenia have deficits in facial emotion recognition and suggested that this is consistent with an endophenotypic process, and that understanding this association further may help with early detection and treatment of the disorder. There have been a number of studies examining the relationship between genetic risk for schizophrenia and facial emotion recognition deficits published since that review, whilst the increasing availability of data from genome-wide association studies (GWAS) opens up the possibility of using individual-level genetic data, rather than family history, as a means for studying the association between schizophrenia genetic risk and facial emotion recognition deficits.

We aimed to systematically review the literature reporting relationships between genetic risk for schizophrenia in unaffected individuals, as indexed either by family history or individual-level genetic data, and facial emotion recognition deficits, and to examine whether this association was stronger for specific emotions.
2 Methods

A systematic review was performed in accordance with PRISMA guidelines (Moher et al., 2009). The full search protocol was pre-registered on PROSPERO (ID: CRD42018088114). Whilst this protocol was developed to also identify studies examining cognitive biases associated with psychosis, we only present results for facial emotion recognition deficits in this paper.

2.1 Literature search
The following databases were searched (by DM) from inception up to October 2017: PsychINFO, MEDLINE, EMBASE, and MEDLINE-in-process. The search terms and strategy are available in our supplementary document. We restricted the search to published, peer reviewed studies in the English language. The reference lists of included studies were hand searched. Authors of conference abstracts without full text papers were contacted to request study data.

2.2 Inclusion and exclusion criteria
Inclusion criteria are detailed in the supplementary materials (protocol and screening checklist). Articles must have been published in a peer-reviewed journal and compared performance on a facial emotion recognition task between participants at higher genetic risk for schizophrenia with those at lower genetic risk. Studies that examined task performance in individuals with a diagnosis of schizophrenia were excluded.

2.3 Definition of high genetic risk for schizophrenia
High genetic risk for schizophrenia was defined as having a higher number of risk alleles for schizophrenia, having more copy number variants associated with schizophrenia, or having one or more first-degree relatives with schizophrenia. Where studies included first-degree of relatives of people with psychotic disorders more broadly, we set a threshold of at least 70% of these having schizophrenia as an inclusion criterion.

2.4 Definition of facial emotion recognition tasks
Facial emotion recognition tasks included any test that measured a participant’s accuracy in identifying the emotional state of another person by the observation of their facial expression. This is typically achieved by showing participants photographs of a variety of people with different facial expressions and identifying the emotion from several response options.

2.5 Data collection

One author (DM) screened all abstracts and obtained full texts of papers that potentially met inclusion criteria. Working independently, two authors (DM and JC) screened full-text articles to determine if they met inclusion criteria (see ‘Screening Checklist’ in Supplementary Materials). Data were extracted independently (by D.M and either AP or DS). Any discrepancies in decisions at any stage of the screening were resolved following discussion with a third reviewer (SZ).

2.6 Quality assessment

The quality of individual studies was assessed by two independent reviewers (DM and either AP or DS) using an assessment checklist which the reviewing team designed based on the Newcastle-Ottawa Scale, a widely used risk of bias tool for observational studies. Total scores, out of maximum of four points, were calculated based on how many of the following criteria each study fulfilled: i) random, consecutive or complete sampling; ii) response rate given; iii) appropriate consideration of confounders (e.g. adjusting for variables that were more plausible as confounders than as mediators); iv) low genetic risk (control) group comparable to high risk group, based on selection method. For studies using genetic data to define level of risk, we assessed whether confounding by ethnicity/population stratification was addressed.

2.7 Data analysis

Where adequate data were provided by study authors, a meta-analysis was performed using the metan command in Stata 15. A random effects model was used due to the differences between methods in the included studies. Random effects models are more conservative than fixed-effects models and generate wider confidence intervals. The test score mean, standard deviation and sample size (n) for both the high and low risk groups were used to
derive a standardised mean difference (SMD) and confidence intervals (CI) for each study. For studies that split high genetic risk participants into separate groups (e.g. siblings and parents), the means and standard deviations were combined according to Cochrane guidelines (Higgins and Green, 2008). Where insufficient data were available to conduct a meta-analysis, studies were summarised using a narrative synthesis. Between-study heterogeneity was estimated using the $I^2$ statistic, and potential reasons for heterogeneity were examined using meta-regression (metareg command in Stata). We examined the following pre-specified variables as potential sources of variation in effect estimates: i) score on our quality assessment tool, ii) whether or not the facial emotion recognition test had a stated time limit (some tests limited the participants to answer within 5 seconds, some had no limit), and iii) whether the test used to assess facial emotion recognition had been previously validated. The likelihood of publication bias was examined using a Funnel Plot and egger test.

3 Results

2927 references were identified in the search after removal of duplicates. After screening against title and abstract, 105 studies were assessed for full text eligibility, and 34 studies were included in this systematic review (see Table 1 for summary of included studies and Figure 1 for PRISMA flow diagram with reasons for exclusion). The included studies were from a range of countries, with the most common (35%) being the USA. The earliest study was in 1989 and the most recent 2017. Study sample sizes ranged from 24 to 4097 (median = 68). There were four studies in children (between 6-15 years old) and one in younger people aged 13-25 years old. The other 30 studies included adults across a wide age range. Two studies used polygenic risk scores informed by genome-wide association studies (GWAS) of schizophrenia to characterise genetic risk in unaffected individuals, and the remaining studies selected individuals based on the presence or absence of a family history of schizophrenia in first-degree relatives.
3.1 Facial emotion recognition tests used
The majority of studies (94%) used a facial emotion “identification” test: where the participants identified the emotion from a list of multiple options. The other tests used were facial emotion “discrimination”: discriminating between only two options; and facial emotion “valence”: choosing between whether the facial emotion is either positive or negative.

3.2 Meta-analyses
We were able to include 23 studies that identified those at high genetic risk of schizophrenia based on family history in a meta-analysis of overall facial recognition score (Figure 2). This indicated strong evidence of a deficit in overall facial emotion recognition among people with a family history of schizophrenia compared with controls (SMD 0.38, 95%CI 0.26 to 0.51, p = <0.001). There was moderate heterogeneity between the included studies with an I² of 41.38% (p<0.007). However, meta-regression showed that none of the variables tested explained this heterogeneity (see supplementary document). There was little evidence of possible publication bias (see Figure 3 for funnel plot; Egger test p = 0.54).

When pooling all negative emotions together (anger, disgust, fear and sadness), meta-analysis (figure 4) shows that there was strong evidence of a deficit in those with a first-degree relative with schizophrenia compared to controls (studies = 25, total n = 3964, SMD 0.21 CI 0.09 to 0.33, p = 0.001; I² = 64.6%). There was no difference for positive valence (happy) facial emotion recognition (figure 5, studies = 6, total n = 588, SMD -0.05, CI -0.31 to 0.22, p = 0.734) or neutral faces (figure 6, studies = 2, total n = 163, SMD -0.01, CI -0.31 to 0.30, p = 0.968, I² = 0%), but these were based on fewer studies and confidence intervals overlapped substantially with those for negative emotions.

We also performed separate meta-analyses on the eight studies that presented results for specific emotions (anger, disgust, fear, sadness, surprise and happiness, see supplementary documents for meta-analyses Forrest plots). There was some evidence that first-degree relatives were worse than controls at recognising anger (studies = 8, total n=1276; SMD 0.27, 95%CI 0.13 to 0.42; p = <0.001; I² = 19.7%) and disgust (studies = 3, total n = 308, SMD
0.37 CI 0.07 to 0.67, p = 0.017, I² = 26.5%). The standardised mean differences were also lower for first-degree relatives compared with controls, but the evidence was much weaker, for fear (studies = 8, total n = 1255, SMD 0.23, CI -0.05 to 0.50, p = 0.103, I² = 72.4%), surprise (studies = 3, total n = 308, SMD 0.24, CI -0.08 to 0.56, p = 0.141, I² = 33.1%) and sadness (studies = 6, total n = 1125, SMD 0.09 CI -0.20 to 0.39, p = 0.55; I² = 71.7%).

3.3 Narrative summaries
In addition to the studies presented in the results section, our review includes three studies of first-degree relatives vs control group that we were unable to incorporate into our meta-analysis. These studies did not provide data that to allow inclusion in our meta-analyses. All three studies showed results in keeping with the results of our meta-analyses. Two provided strong evidence that those at high risk of schizophrenia performed less well at overall facial emotion recognition than those at low risk (Cohen’s d -0.31, p<0.001 (Kohler et al., 2014); Cohen’s d -0.39, p<0.0001 (Calkins et al., 2010). In the third study (Yang et al., 2015) separate emotions were tested, and we calculated p-values comparing relatives to controls based on the means for both high and low intensity of each emotion presented in the paper. The strongest evidence of a deficit in the relatives group compared to control group was p = 0.068 for recognition of high intensity fear. We were unable to combine the results for each intensity of emotion to allow us to include this study in our meta-analysis because the proportion of faces at each intensity was not stated.

3.4 Studies based on polygenic risk scores
Two studies used polygenic risk scores to define genetic risk, the first examining multiple risk scores derived using different p-thresholds (pT), and the second examining a single risk score derived at a pT of <0.05. In a study of facial emotion recognition ability in 8-year-old children within the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort (n = 4097), there was weak evidence of poorer performance on facial emotion recognition tasks in those with higher genetic risk when testing for sad (p=0.032), but not happy (p = 0.222), angry (p = 0.175), or fearful (p = 0.456) emotions (Coleman et al., 2017). In the other study, of 4303 participants aged 8-21 in the USA, there was no evidence that polygenic risk for schizophrenia was associated with facial emotion recognition accuracy (Germine et al., 2016).
3.5 Quality assessment

The quality assessment of included studies is shown in supplementary documents, table 1. In summary, only two of the 34 included studies (6%) reported a participation response rate. Three studies (9%) reported using random, consecutive or complete sampling. Seven (21%) had a low genetic risk group that was deemed comparable to the high genetic risk group based on the sampling strategy. 6 studies (18%) adjusted for variables that we considered could be confounding factors of the association between genetic risk for schizophrenia and cognitive biases. 16 (47%) adjusted for variables that are more plausible as mediators than confounders of the association.

4 Discussion

This review presents a summary of the findings from our systematic review of the research examining the relationship between genetic risk of schizophrenia and facial emotion recognition deficits. We are able to update the research of this relationship since the meta-analysis performed by Lavoie et al in 2012 by including 17 additional studies investigating overall emotion recognition deficits. We also present additional meta-analyses of facial emotion recognition deficits for negative, positive and neutral valence, as well as specific facial emotions including anger, disgust, fear, sadness, surprise and happiness. Our findings demonstrate strong evidence of deficits in negative facial emotion recognition in those with first-degree relatives with schizophrenia, but no evidence of a deficit for recognition of neutral or positive facial emotion recognition. All three studies included as narrative summarises in the results section showed results in keeping with the results of our meta-analyses.

Test scores for specific emotions showed strong evidence for a deficit in recognising anger and disgust amongst first-degree relatives of people with schizophrenia, weaker evidence of deficits in recognition for fear and surprise, and no evidence of a difference for recognising happy or neutral faces. Our meta-analysis showed that there was strong evidence of overall deficits in recognising facial emotions with a negative valence in people with first-degree
relatives with schizophrenia. Such a deficit in interpreting other people’s negative emotional states could potentially lead to misinterpretation of situations. However, the majority of studies included in our review did not present data for specific facial emotion recognition deficits, so we are unable to draw firm conclusions as to whether there are deficits in specific emotions in relatives of people with schizophrenia, or whether deficits are only for negative emotions give that few studies examined positive or neutral emotions.

Our review also includes two studies that examined an association between polygenic risk scores for schizophrenia and facial emotion recognition test scores. Neither study reported clear evidence of an association, although one reported an association with facial emotion recognition speed. As there was no association with sensorimotor speed, the authors hypothesised that speed of facial emotion recognition might slow before deficits in recognition are apparent. Given the substantially larger sample sizes of the two polygenic risk studies (mean N = 4200) compared with the family history studies (mean N = 192), it is somewhat surprising that the evidence of facial emotion recognition deficits was so much weaker in the former, given the likely increased statistical power of these studies. It is possible that change in facial emotion recognition ability does not occur across the continuum of genetic risk, but only at the high risk end (which would be more likely to be captured by sampling first-degree relatives, hence offsetting the power loss due to smaller sample sizes). The studies that examined genetic risk scores did not test non-linear models to explicitly test this hypothesis. Another explanation is that genetic risk is not causally related to facial emotion recognition, and the association with family history of schizophrenia is confounded by other characteristics related to family environment, such as increased levels of stress or adversity in children where a parent or sibling has schizophrenia.

Our quality assessment found that the included studies were generally of poor quality. In our meta-analysis, we found strong evidence of an associated between emotion recognition deficits in those with first-degree relatives with schizophrenia compared with those without. However, there was moderate heterogeneity between studies which was not explained by our meta-regression of study quality, test used or whether test response was time limited.
An important observation in our review is that less than half of the studies (44%) included made an attempt to address confounding. However, most adjusted for characteristics such as educational attainment, IQ, other measures of cognitive function, substance use, and psychiatric symptoms, which are perhaps easier to envisage as potential mediators of the effect of family history (or genetic risk for schizophrenia) on emotion recognition, rather than as confounders of this relationship. If this is the case, then adjusting for these would lead to an underestimate of the true causal effect of genetic risk on cognitive biases in these studies.

Our results suggest that facial emotion recognition deficits are not a consequence of schizophrenia given that these deficits are present in unaffected first-degree relatives. The effect size of these deficits was always lower in first-degree relatives than those reported in people with schizophrenia (Alfimova et al., 2013; Bediou et al., 2007), which might reflect a greater genetic risk, although could also indicate that pre-morbid deficits increase in some people with schizophrenia following the first episode of psychosis. More longitudinal research is required to determine if emotion recognition deficits increase the risk of the development of schizophrenia onset or relapse. However, such observational studies will always be limited in their ability to determine causality due to concerns around residual confounding. It is probably only through trials of interventions targeting emotion recognition deficits or causal inference methods, such as Mendelian randomisation (once genetic instruments for these deficits become available, that our ability to determine causal effects of emotion recognition on schizophrenia will be substantially improved.

Computational and animal models of perception and learning (Fletcher and Frith, 2009) implicate dopaminergic and glutamatergic function as fundamental pathways involved in both perception and belief formation, whilst genes involved in these pathways have been identified as risk variants for schizophrenia in a recent GWAS (Schizophrenia Working Group of the Psychiatric Genomics, 2014). Furthermore, genetic risk for schizophrenia has been associated with greater risk of being exposed to childhood trauma, which itself has been associated with facial emotion recognition deficits (da Silva Ferreira et al., 2014). Plausible explanations of how higher genetic risk for schizophrenia could lead to deficits in facial
emotion recognition, therefore exist, although given the limitations of our review, stronger evidence is required that this association is causal and not due to bias or confounding.

4.1 Strength and limitations of review
We were able to include a high number of studies which compared facial emotion recognition in those at high genetic risk of schizophrenia with those at low risk and this included two studies which utilised the polygenic risk score approach using data from a recent schizophrenia GWAS. We followed PRISMA guidelines throughout the review (see supplementary document). The meta-analysis was based on data from a large number of participants (n=3947) and the majority of studies used tests for facial emotion recognition that could be standardised in pooled analysis.

There are also a number of important limitations with our review. Although we carried out a systematic and thorough search and review of the peer-reviewed, published literature, we may nevertheless have missed some studies that could have contributed to addressing our study aims, particularly given our restriction of only including English-language publications. We were also unable to include all studies in the meta-analyses as some studies did not provide the data required in the paper or on request. The lack of clear information in the methods section of some included studies also made precise exploration of the quality of studies and differences between studies difficult. This highlights the importance for authors to include all results in numerical form, and of a thorough documentation of study methods. The conclusions drawn from any review is reliant on the quality of the studies included, and our quality assessment shows that the included studies were generally of poor quality, making it difficult to draw firm conclusions about the role of schizophrenia genetic risk on the facial emotion recognition deficits explored. Finally, as we carried out multiple meta-analyses, we are cautious in the interpretation of our results, particularly for the valence and emotion specific ones, as these were not our primary exposure of interest.
4.2 Conclusions

Studies using family history as a marker for genetic risk need to carefully consider the potential effects of confounding and distinguish this from mediation to allow appropriate inferences about causal effects to be made. Availability of molecular genetic data to use polygenic risk scoring and Mendelian randomisation (Davies et al., 2018) approaches could help address issues of confounding and causal direction in future studies and thereby help to clarify whether genetic risk for schizophrenia has a causal effect on facial emotion recognition deficits.


Table 1: Summary of included studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Size</th>
<th>high risk N</th>
<th>mean age</th>
<th>% male</th>
<th>control group N</th>
<th>mean age</th>
<th>% male</th>
<th>Measure</th>
<th>High risk group Mean</th>
<th>SD</th>
<th>Control group Mean</th>
<th>SD</th>
<th>Confounders adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albacete 2019</td>
<td>Spain</td>
<td>74</td>
<td>37</td>
<td>48.8</td>
<td>43.2</td>
<td>37</td>
<td>41.07</td>
<td>48.6</td>
<td>FEIT</td>
<td>101.3</td>
<td>17.86</td>
<td>110.7</td>
<td>11.37</td>
<td>None</td>
</tr>
<tr>
<td>Alfimova 2009</td>
<td>Russia</td>
<td>154</td>
<td>55</td>
<td>45</td>
<td>50.9</td>
<td>99</td>
<td>32</td>
<td>33.3</td>
<td>FEIT</td>
<td>5.25</td>
<td>5.46</td>
<td>5.8</td>
<td>1.2</td>
<td>Sex, age, education</td>
</tr>
<tr>
<td>Allott 2015</td>
<td>Austria</td>
<td>57</td>
<td>27</td>
<td>32.48</td>
<td>37</td>
<td>30</td>
<td>15.64</td>
<td>50</td>
<td>FEIT</td>
<td>Results as separate emotions - see supp.</td>
<td>Age, IQ, symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Andersen 2016</td>
<td>USA</td>
<td>75</td>
<td>28</td>
<td>30.7</td>
<td>17.9</td>
<td>47</td>
<td>26.8</td>
<td>53.2</td>
<td>FEIT</td>
<td>13</td>
<td>2.4</td>
<td>13</td>
<td>2.9</td>
<td>Age, sex</td>
</tr>
<tr>
<td>Andric 2016</td>
<td>Serbia</td>
<td>106</td>
<td>55</td>
<td>28.56</td>
<td>41.8</td>
<td>51</td>
<td>29.8</td>
<td>45.1</td>
<td>FEIT</td>
<td>76.94</td>
<td>8.41</td>
<td>78.1</td>
<td>9.6</td>
<td>Age, gender, IQ, general facial recognition</td>
</tr>
<tr>
<td>Ay 2016</td>
<td>Turkey</td>
<td>60</td>
<td>30</td>
<td>42.4</td>
<td>30</td>
<td>30</td>
<td>33.13</td>
<td>60</td>
<td>FEIT</td>
<td>12.56</td>
<td>1.99</td>
<td>14.2</td>
<td>1.88</td>
<td>None</td>
</tr>
<tr>
<td>Bediou 2007</td>
<td>Morocco</td>
<td>56</td>
<td>30</td>
<td>31.2</td>
<td>100</td>
<td>26</td>
<td>24.3</td>
<td>100</td>
<td>FEIT</td>
<td>0.565</td>
<td>0.10</td>
<td>0.6</td>
<td>0.10</td>
<td>Age, education</td>
</tr>
<tr>
<td>Bolte 2003</td>
<td>Germany</td>
<td>66</td>
<td>46</td>
<td>41.30</td>
<td>49.99</td>
<td>22</td>
<td>29.7</td>
<td>50</td>
<td>FEIT</td>
<td>38.94</td>
<td>4.39</td>
<td>42.9</td>
<td>1.3</td>
<td>Age, IQ</td>
</tr>
<tr>
<td>Calkins 2010</td>
<td>USA</td>
<td>1262</td>
<td>928</td>
<td>45.6</td>
<td>32.21</td>
<td>334</td>
<td>42.4</td>
<td>42.8</td>
<td>FEIT</td>
<td>Results given as cohen d = -0.39 p&lt;0.0001</td>
<td>Sex, age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cella 2015</td>
<td>UK</td>
<td>42</td>
<td>21</td>
<td>33.7</td>
<td>52.4</td>
<td>21</td>
<td>27.3</td>
<td>28.6</td>
<td>FEIT</td>
<td>13.52</td>
<td>1.8</td>
<td>14.5</td>
<td>1.2</td>
<td>Age, gender, education, IQ, cognitive function</td>
</tr>
<tr>
<td>Coleman 2017</td>
<td>UK</td>
<td>4097</td>
<td>4097</td>
<td>8</td>
<td>49.6</td>
<td>4097</td>
<td>8</td>
<td>49.6</td>
<td>FEIT</td>
<td>Results as separate emotions - see supp.</td>
<td>White western European ancestry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davalos 2004</td>
<td>USA</td>
<td>102</td>
<td>51</td>
<td>10.22</td>
<td>64.7</td>
<td>51</td>
<td>10.45</td>
<td>64.7</td>
<td>FEIT</td>
<td>18.85</td>
<td>8.38</td>
<td>17.8</td>
<td>2.7</td>
<td>None</td>
</tr>
<tr>
<td>deAchaval 2010</td>
<td>Argentina</td>
<td>40</td>
<td>20</td>
<td>50.1</td>
<td>45</td>
<td>20</td>
<td>44.2</td>
<td>45</td>
<td>FEIT</td>
<td>16.6</td>
<td>2.8</td>
<td>18.1</td>
<td>1.5</td>
<td>Age, cognitive performance, education</td>
</tr>
<tr>
<td>Erol 2010</td>
<td>Turkey</td>
<td>116</td>
<td>58</td>
<td>34.4</td>
<td>58.6</td>
<td>58</td>
<td>33.6</td>
<td>60.3</td>
<td>FEIT</td>
<td>12.2</td>
<td>2.5</td>
<td>13.7</td>
<td>1.6</td>
<td>Age, education, illness duration and BPRS</td>
</tr>
<tr>
<td>Germine 2016</td>
<td>USA</td>
<td>4303</td>
<td>4303</td>
<td>13.8</td>
<td>50</td>
<td>4303</td>
<td>13.8</td>
<td>50</td>
<td>FEIT</td>
<td>Numerical data not presented</td>
<td>White non-Hispanic ancestry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goghari 2011</td>
<td>USA</td>
<td>59</td>
<td>23</td>
<td>49.8</td>
<td>35</td>
<td>36</td>
<td>42.1</td>
<td>66</td>
<td>FEIT</td>
<td>Results as separate emotions - see supp.</td>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goghari 2017</td>
<td>Canada</td>
<td>46</td>
<td>25</td>
<td>41.2</td>
<td>40</td>
<td>21</td>
<td>43.4</td>
<td>52.4</td>
<td>FEIT</td>
<td>Results as separate emotions - see supp.</td>
<td>None</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goldschmidt 2014</td>
<td>Argentina</td>
<td>28</td>
<td>14</td>
<td>30.4</td>
<td>57</td>
<td>14</td>
<td>28.4</td>
<td>57</td>
<td>FEIT</td>
<td>97</td>
<td>5</td>
<td>99</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>Horton 2017</td>
<td>USA</td>
<td>50</td>
<td>16</td>
<td>13.6</td>
<td>37</td>
<td>34</td>
<td>11.9</td>
<td>41</td>
<td>FEIT</td>
<td>57.81</td>
<td>8.69</td>
<td>62.94</td>
<td>10.14</td>
<td>Age</td>
</tr>
<tr>
<td>Huepe 2012</td>
<td>Not stated</td>
<td>32</td>
<td>14</td>
<td>45.79</td>
<td>40</td>
<td>18</td>
<td>40.5</td>
<td>61.1</td>
<td>FEIT</td>
<td>-0.99</td>
<td>0.6</td>
<td>-1.05</td>
<td>0.6</td>
<td>None</td>
</tr>
<tr>
<td>Ibanez 2012</td>
<td>Not stated</td>
<td>26</td>
<td>13</td>
<td>47.31</td>
<td>46.15</td>
<td>13</td>
<td>39.48</td>
<td>69.23</td>
<td>FEVT</td>
<td>76</td>
<td>6.93</td>
<td>81.5</td>
<td>6.36</td>
<td>None</td>
</tr>
<tr>
<td>Kee 2004</td>
<td>USA</td>
<td>100</td>
<td>51</td>
<td>38.64</td>
<td>43</td>
<td>49</td>
<td>36.38</td>
<td>49</td>
<td>FEIT</td>
<td>13.41</td>
<td>2.52</td>
<td>14.1</td>
<td>2.39</td>
<td>None</td>
</tr>
<tr>
<td>Kohler 2014</td>
<td>USA</td>
<td>142</td>
<td>52</td>
<td>20.4</td>
<td>52</td>
<td>90</td>
<td>21.3</td>
<td>51.11</td>
<td>FEIT</td>
<td>Numerical data not presented</td>
<td>None</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lavoie 2014</td>
<td>Canada</td>
<td>69</td>
<td>31</td>
<td>56.1</td>
<td>29</td>
<td>38</td>
<td>54.55</td>
<td>29</td>
<td>FEIT</td>
<td>11.42</td>
<td>1.43</td>
<td>11.84</td>
<td>1.42</td>
<td>IQ, non-social reasoning, age, gender, SES</td>
</tr>
<tr>
<td>Leppanen 2008</td>
<td>South Africa</td>
<td>45</td>
<td>23</td>
<td>36</td>
<td>35</td>
<td>22</td>
<td>40.9</td>
<td>50</td>
<td>FEIT</td>
<td>Results as separate emotions - see supp.</td>
<td>None</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Li 2010</td>
<td>China</td>
<td>90</td>
<td>23</td>
<td>30</td>
<td>52</td>
<td>67</td>
<td>26</td>
<td>67</td>
<td>FEIT</td>
<td>57.93</td>
<td>21.84</td>
<td>64.3</td>
<td>11.32</td>
<td>None</td>
</tr>
<tr>
<td>Li 2012</td>
<td>China</td>
<td>24</td>
<td>12</td>
<td>31.25</td>
<td>33.3</td>
<td>12</td>
<td>29.25</td>
<td>50</td>
<td>FEIT</td>
<td>Results as separate emotions - see supp.</td>
<td>None</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McCown 1989</td>
<td>USA</td>
<td>100</td>
<td>50</td>
<td>45.31</td>
<td>50</td>
<td>50</td>
<td>46.45</td>
<td>50</td>
<td>FEIT</td>
<td>70.04</td>
<td>7.97</td>
<td>73.7</td>
<td>9.34</td>
<td>None</td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Sample Size</td>
<td>Mean Age</td>
<td>Female</td>
<td>Male</td>
<td>FEIT</td>
<td>Results Note</td>
<td>Age, Gender, Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>----------</td>
<td>-------------</td>
<td>----------</td>
<td>--------</td>
<td>------</td>
<td>------</td>
<td>------------------------------------------------</td>
<td>--------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mendoza 2011</td>
<td>Cuba</td>
<td>219</td>
<td>42.01*</td>
<td>59.1</td>
<td>109</td>
<td>36.7</td>
<td>Results as separate emotions - see supp.</td>
<td>Age, gender, education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ruocco 2014</td>
<td>USA</td>
<td>712</td>
<td>42.65</td>
<td>30</td>
<td>380</td>
<td>47</td>
<td>ER-40</td>
<td>0.34 0.07 0 1 Age, race and sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spilka 2017</td>
<td>Canada</td>
<td>54</td>
<td>41.19</td>
<td>37</td>
<td>27</td>
<td>4.1</td>
<td>FEDT</td>
<td>90.65 4.72 89.73 4.03 None</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toomey 1999</td>
<td>USA</td>
<td>40</td>
<td>46.2</td>
<td>24</td>
<td>19</td>
<td>21</td>
<td>FEIT</td>
<td>55.3 7.1 57 3.8 None</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wolf 2011</td>
<td>USA</td>
<td>45</td>
<td>42.3</td>
<td>45</td>
<td>25</td>
<td>48</td>
<td>FEIT</td>
<td>91.5 7.2 94.4 4.6 None</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yang 2015</td>
<td>China</td>
<td>56</td>
<td>23.9</td>
<td>46.2</td>
<td>30</td>
<td>24.6</td>
<td>FEIT</td>
<td>Results separated by emotion and intensity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Notes on table 1: Highlighted studies were included in meta-analysis by Lavoie et al (2013). * median values. FEIT facial emotion identification test. FEDT facial emotion discrimination test. FEVT facial emotion valence test. BT beads task. IPSAW Internal, personal and situations attributions questionnaire.

Figure 1
Prisma Flow Diagram
(see file figure 1)

Figure 2
Forrest Plot for Meta-analysis of standardised mean difference for facial emotion recognition test
(see file figure 2)

Footnote:
Markers signify which studies are additional inclusions to the studies reviewed in the paper by Lavoie et al 2013.

Positive results demonstrate higher standardised mean difference between scores on facial emotion recognition tests achieved by people at low genetic risk of psychosis compared with those at high genetic risk of psychosis.
Figure 3
Funnel plot for meta-analysis of facial emotion recognition
(See file figure 3)

Figure 4
Forrest Plot for Meta-analysis of standardised mean difference for facial emotion recognition test for facial expressions with negative valence.
(see file figure 4)
Footnote:
Positive results demonstrate higher standardised mean difference between scores on facial emotion recognition tests achieved by people at low genetic risk of psychosis compared with those at high genetic risk of psychosis.

Figure 5
Forrest Plot for Meta-analysis of standardised mean difference for facial emotion recognition test for facial expression with positive valence
(see file figure 5)
Footnote:
Positive results demonstrate higher standardised mean difference between scores on facial emotion recognition tests achieved by people at low genetic risk of psychosis compared with those at high genetic risk of psychosis.
Source of funding

This project was not directly funded.
Acknowledgments

This study was supported by the NIHR Biomedical Research Centre at University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health and Social Care.
Conflict of Interest

None of the authors have any conflicts of interest in relation to this work.
Contributors
All authors have contributed substantially towards the design of the study, the analysis and interpretation of the data, and drafting the manuscript. All authors have approved the final version. David Martin had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Systematic review and meta-analysis of the relationship between genetic risk for schizophrenia and facial emotion recognition

Reviewers' Comments and author responses

Reviewer 1 comments:

1. A main concern is what the current findings add above and beyond the meta-analysis conducted by Lavoie et al. (2013), other than the addition of some additional articles added to the meta-analysis. The authors report that using individual-level data from GWAS studies may improve our ability to assess associations with genetic risk in emotion recognition deficits, yet review is only provided for 2 studies, which ultimately do not support this position. I believe the study would be more impactful if additional data could be provided. E.g., Meta-analysis of negative (combining anger, fear, disgust) vs. positively valanced (happy) vs. neutral emotions for those studies that report emotion specific deficits.

   We appreciate this concern about what this review adds above and beyond the meta-analysis by Lavoie et al (2013) other than updating the review through the addition of 17 studies that we have included that were not in that review. As suggested, we have now also included meta-analyses of negative (anger, fear, disgust and sadness), positive and neutral emotions for those studies which report specific deficits. This was a useful addition to our manuscript as it highlights how strong the evidence is for deficits in the recognition of negative emotions in those with first-degree relatives of people with schizophrenia. We have expanded our description of what our reviews adds to this area of research in the discussion section (first paragraph).

2. One study that is omitted from the meta-analysis is that by Ruocco et al, 2014 from the B-SNIP consortium. This is a large study (n>250 schizophrenia relatives) using a well-validated emotion recognition task, and based on the selection criteria it is not clear why this study is not included. This seems particularly important to include, as it is among the larger studies reported in the literature and it was published after the Lavoie et al. 2013 meta-analysis. Further, data are reported for specific emotions, which would allow for inclusion in the meta-analysis on the subset of studies that evaluation specificity of deficits.

   We are grateful to the reviewer for pointing out this omission. This study was picked up in our search but was mistakenly omitted from our review, for which we do apologise. We have now added this study to our review and included it in all meta-analyses, and it has proved to be influential in strengthening the findings of our review.

3. Through the manuscript (e.g., 2nd sentence of abstract, 1st sentence of 3rd paragraph of introduction, etc.) the authors use language specifying a "causal effect" of emotion recognition deficits on schizophrenia, which I believe overstates what the studies they review are reasonably measuring, which are really associations. I appreciate the authors point that studying unaffected relatives may remove some of the confounds and cause vs. effect issues when studying affected individuals, but I think tempering of this language would improve the manuscript.

   We had not meant to give the impression that the evidence of association between emotion recognition deficits and schizophrenia is causal, but only that studies that can minimize other explanations such as reverse causation and confounding are required to more robustly understand this relationship. We have therefore amended the text in the abstract and main text to tone down any suggestion that the association is necessarily causal.

4. Is any statistical correction for multiple meta-analyses provided for the evaluation of deficits in specific emotions?

   We did look into whether any statistical correction for multiple meta-analyses was required, and decided to follow the advice from Cochrane systematic review guidance (see link below), which states that statistical correction for meta-analyses is not generally recommended. However, Cochrane do advise that overall conclusions are more difficult to draw if there are multiple analyses. We have therefore encouraged cautiousness in interpretation of our results in the discussion section (section 4.1, second paragraph)

https://handbook-5-1.cochrane.org/chapter_16/16_7_2_multiplicity_in_systematic_reviews.htm
5. The authors should report effect sizes for the findings from reviews of the studies using polygenic risk scores vs. p values alone. As the authors note in the discussion, the substantial subject sizes of these studies increase statistical power, which indicates finding of statistical significance (p=.032) reflect a very small effect.

We agree that this would be useful, but unfortunately the studies haven't standardised their scores and so effect sizes are not interpretable, as outlined by Mistry et al (https://www.ncbi.nlm.nih.gov/pubmed/29129507).

Reviewer #3:

The authors mention that a systematic review and meta-analysis conducted in this topic was reported in 2013. In this context, the authors have mentioned that 16 additional studies have been included in this present meta-analysis that was not evaluated in the previous one. In the figure 2 that has listed these studies, many are found to be published 2012 - it will be helpful if the authors could list the 16 additional studies separately.

We have now added markers on figure 2 to signify which studies are additional inclusions to the studies reviewed in the paper by Lavoie et al 2013.

The authors may want to explain the additional new findings that are available through this present meta-analysis.

We have added more detail in our discussion section (first paragraph) on the new findings available through our meta-analyses; namely, an update of the evidence through the inclusion of 17 additional studies in our review, 11 additional studies in our overall facial emotion recognition meta-analysis, and our meta-analyses for separate valences and specific emotions.

The supplementary material document could not be accessed by this reviewer to learn about the details of the excluded studies. For example, fMRI study by van der Velde et al 2015 may be a potential research publication for this meta-analysis. Similar studies that have evaluated emotion processing with concurrent assessment of a neurobiological parameter (like fMRI) are important to consider.

Thank you for this suggestion and we are sorry to hear that you were unable to see our supplementary documents. The study by Velde et al from 2015 does not meet our inclusion criteria as they recruited people at "ultra high-risk" based on CAARMS scoring rather than on genetic (or familial) risk. In reference to the inclusion of studies assessing neurobiological parameters, we have included a number of such studies where they met our inclusion criteria (for example: Goghari et al 2017, Spilka et al 2017 and Wolf et al 2011), and extracted the relevant data for the purpose of this review.

With regards to emotion processing task, it will be helpful to know the summary of different type of experiments (for example using static versus dynamic stimuli / similar other parameters) and the potential impact of these parameters on the findings.

All included studies used static stimuli, some with a time-limit for how long the stimuli were shown and some with no time limits. We examined whether differences in the use of time limits reduced heterogeneity as part of our meta-regression work (section 3.2, first paragraph and in supplementary document).

Reviewer #4:

1. The main concern with this manuscript is related to the conclusion that the authors are drawing based on the results of this meta-analysis. It is unclear if showing that first-degree relatives with schizophrenia have lower emotion recognition is in any way indicative that this endophenotype could not also be partly a consequence of schizophrenia. First-degree relatives of schizophrenia have about 10 times more risk of developing the illness compared to the general population. More longitudinal research in youth at genetic risk of schizophrenia is needed to determine whether lower emotion recognition could increase the risk of developing the illness. Furthermore, it is possible that some people with schizophrenia develop more pronounced emotion recognition deficits following the first episode of psychosis. These last points should probably be address in the discussion, and I would advise to remove the references to causality (or the absence of) in the manuscript.
We believe that reverse causation (schizophrenia leading to emotion recognition deficits) is an unlikely explanation for our findings as participants in the majority of studies included in our review had been screened to ensure they did not have schizophrenia. However, we agree with the reviewer that further longitudinal research is required to determine if emotion recognition deficits have a causal effect on schizophrenia onset or relapse. We also appreciate that such observational studies will always be limited in their ability to determine causality due to concerns around residual confounding. It is probably only through trials of interventions that target such deficits, or causal inference methods such as Mendelian randomisation (once genetic instruments for these deficits become available) that our ability to determine causal effects of emotion recognition on schizophrenia will be substantially improved. We have added these points to the discussion section (Paragraph 5) and re-phrased the sections referring to causality (introduction and discussion).

2. In this meta-analysis, the authors included individuals with first-degree relatives with psychosis. The authors noted that only 70% of these individuals had schizophrenia (mentioned in section 2.3). What was the diagnosis for the other 30%? How did the authors ensure that these studies fall within the inclusion criteria for genetically high-risk for schizophrenia? The differences between schizophrenia related risk and/or non-schizophrenia risk should be further explained for each individual study included in the analysis.

Thank you for raising this important point. This criterion was actually introduced in relation to a separate systematic review we were conducting in parallel that was examining other cognitive biases in first-degree relatives, and should not have been applied to this review where our original protocol specified that all participants should have had a first-degree relative with schizophrenia. There were 2 studies that included participants who had a first-degree relative with other, non-schizophrenia psychotic disorders in our original submission, and these have now been excluded to ensure that all included studies met our original protocol.

3. The current Table 1 on the quality of assessment tool could be moved into supplementary material. In replace of Table 1 within the manuscript, the authors could include a Table 1 that is more informative of the studies included in the meta-analysis in their manuscript. For example, there could be a Table presenting the main characteristics of each study included in the meta-analysis (e.g., n=, diagnosis of first-degree relatives, age, sex ratio, IQ, etc.).

We are happy to include a table summarising the included studies main characteristics and therefore we have changed Table 1, as suggested. However, we are aware it is a very large table and therefore leave it to the editorial team to decide where to include this.

4. In 2.6, the authors mention that any discrepancies in quality assessment rating were resolved by discussion with a third reviewer. How was this conducted, and was the process by which the confounding’s were assessed?

Discrepancies in quality assessment rating were resolved after discussion with a third reviewer (SZ) at a meeting where relevant information from the study was assessed to enable a collaborative agreement on what level of criteria the study had met for the discrepancy in question. Our decision on what we deemed to be appropriate consideration of confounders was based on discussions between the reviewing team as to which factors we believed were most plausible as confounding factors based on current knowledge from the literature.

5. How were the meta-analysis and narrative synthesis compared? The inclusion of both in the same manuscript is a bit confusing. Maybe the authors could only focus on the meta-analysis in their results section and discuss the other manuscripts in the discussion?

We believe that these studies should be included in the results section as they were identified in our search strategy and as they still contribute to the body of evidence, even if not to the meta-analysis estimate. As we discuss, the results from the narrative summaries are consistent with the results from the meta-analyses, adding further support to the evidence of association. We have added a sentence to the discussion section to highlight this (first paragraph of discussion, last sentence).

6. Evidence of some publication bias (3.2) was mentioned. How was this accounted for in the study? There is no mention of this process.

We examined publication bias using a funnel plot and egger test, which shows little evidence (p = 0.54) of such bias. We have added further details as to how the methods used in our review attempts to address publication bias to the first paragraph of our results section.
Minor Concerns:

1. In the second paragraph of the introduction, it is stated that facial emotion recognition deficits in schizophrenia are associated with neuro-radiological differences seen on fMRI. What are these differences? Please elaborate.

We have elaborated on these differences in the introduction section (second paragraph), as follows:

“Facial emotion recognition deficits in schizophrenia are associated with an extensive pattern of activation abnormalities on fMRI, consistent with hypoactive emotion recognition networks (Jani and Kasparek, 2018). Compensatory over-activation in the medial prefrontal cortex (MPFC) during threatening faces processing has also been demonstrated (Dong et al., 2017).”

2. It was a bit unclear whether there was a specific genetic link that was assessed in some studies (i.e., specific genome sequence for schizophrenia or heritability). For the two GWAS included, what were the common genes for schizophrenia risk? Please elaborate.

Both included studies used polygenic risk scores derived using the weighted sum of risk alleles (at multiple p-thresholds in one study, and at a p-threshold of <0.05 in the other study) for schizophrenia identified from the GWAS by the Schizophrenia Working Group of the Psychiatric Genomics, 2014. Hence these scores would have been based on thousands of SNPs across the genome. We have added a sentence to the results section to clarify this (section.3.4).

3. In 2.6, the authors mentioned that two independent viewers assessed the quality of the individual studies, but there seems to be three reviewers (DM, AP and DS). Please clarify.

We have clarified the process in section 2.6: (each study assessed by DM and either AP or DS).
Funnel plot with pseudo 95% confidence limits
<table>
<thead>
<tr>
<th>Study</th>
<th>SMD (95% CI)</th>
<th>% Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allott 2015 sad</td>
<td>0.44 (-0.09, 0.96)</td>
<td>3.27</td>
</tr>
<tr>
<td>Goghari 2011 sad</td>
<td>-0.36 (-0.89, 0.16)</td>
<td>3.26</td>
</tr>
<tr>
<td>Goghari 2017 sad</td>
<td>-0.19 (-0.77, 0.39)</td>
<td>2.89</td>
</tr>
<tr>
<td>Huepe 2012 sad</td>
<td>-0.12 (-0.82, 0.58)</td>
<td>2.24</td>
</tr>
<tr>
<td>Mendoza 2011 sad</td>
<td>0.54 (0.27, 0.81)</td>
<td>5.90</td>
</tr>
<tr>
<td>Ruocco 2014 sad</td>
<td>0.00 (-0.15, 0.15)</td>
<td>7.39</td>
</tr>
<tr>
<td>Allott 2015 anger</td>
<td>-0.15 (-0.67, 0.37)</td>
<td>3.31</td>
</tr>
<tr>
<td>Andric 2016 anger</td>
<td>0.29 (-0.09, 0.68)</td>
<td>4.57</td>
</tr>
<tr>
<td>Goghari 2011 anger</td>
<td>-0.17 (-0.69, 0.36)</td>
<td>3.29</td>
</tr>
<tr>
<td>Goghari 2017 anger</td>
<td>0.35 (-0.23, 0.94)</td>
<td>2.86</td>
</tr>
<tr>
<td>Huepe 2012 anger</td>
<td>0.00 (-0.70, 0.70)</td>
<td>2.24</td>
</tr>
<tr>
<td>Leppanen 2008 anger</td>
<td>0.37 (-0.22, 0.95)</td>
<td>2.83</td>
</tr>
<tr>
<td>Mendoza 2011 anger</td>
<td>0.29 (0.03, 0.56)</td>
<td>5.94</td>
</tr>
<tr>
<td>Ruocco 2014 anger</td>
<td>0.41 (0.26, 0.56)</td>
<td>7.37</td>
</tr>
<tr>
<td>Allott 2015 fear</td>
<td>0.82 (0.28, 1.37)</td>
<td>3.15</td>
</tr>
<tr>
<td>Andric 2016 fear</td>
<td>0.17 (-0.21, 0.55)</td>
<td>4.58</td>
</tr>
<tr>
<td>Goghari 2011 fear</td>
<td>-0.45 (-0.98, 0.08)</td>
<td>3.25</td>
</tr>
<tr>
<td>Goghari 2017 fear</td>
<td>0.04 (-0.54, 0.62)</td>
<td>2.89</td>
</tr>
<tr>
<td>Huepe 2012 fear</td>
<td>0.39 (-0.31, 1.10)</td>
<td>2.21</td>
</tr>
<tr>
<td>Li 2012 fear</td>
<td>-0.05 (-0.85, 0.75)</td>
<td>1.82</td>
</tr>
<tr>
<td>Mendoza 2011 fear</td>
<td>0.66 (0.39, 0.93)</td>
<td>5.87</td>
</tr>
<tr>
<td>Ruocco 2014 fear</td>
<td>0.08 (-0.07, 0.23)</td>
<td>7.39</td>
</tr>
<tr>
<td>Allott 2015 disgust</td>
<td>0.01 (-0.42, 0.62)</td>
<td>3.32</td>
</tr>
<tr>
<td>Huepe 2012 disgust</td>
<td>0.16 (-0.54, 0.86)</td>
<td>2.24</td>
</tr>
<tr>
<td>Mendoza 2011 disgust</td>
<td>0.54 (0.27, 0.81)</td>
<td>5.90</td>
</tr>
<tr>
<td>Overall (I-squared = 62.8%, p = 0.000)</td>
<td>0.21 (0.09, 0.33)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

NOTE: Weights are from random effects analysis.
Figure 5

Click here to download high resolution image

<table>
<thead>
<tr>
<th>Study</th>
<th>SMD (95% CI)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andric 2016</td>
<td>-0.23 (-0.61, 0.16)</td>
<td>14.51</td>
</tr>
<tr>
<td>Goghari 2011</td>
<td>-0.22 (-0.75, 0.30)</td>
<td>9.66</td>
</tr>
<tr>
<td>Goghari 2017</td>
<td>0.02 (-0.56, 0.60)</td>
<td>8.34</td>
</tr>
<tr>
<td>Huepe 2012</td>
<td>0.59 (-0.13, 1.30)</td>
<td>6.01</td>
</tr>
<tr>
<td>Li 2012</td>
<td>-0.57 (-1.39, 0.24)</td>
<td>4.79</td>
</tr>
<tr>
<td>Leppanen 2008</td>
<td>-0.37 (-0.96, 0.22)</td>
<td>8.14</td>
</tr>
<tr>
<td>Mendoza 2011</td>
<td>0.25 (-0.01, 0.52)</td>
<td>20.51</td>
</tr>
<tr>
<td>Ruocco 2014</td>
<td>0.14 (-0.01, 0.26)</td>
<td>26.05</td>
</tr>
<tr>
<td>Allott 2015</td>
<td>(Excluded)</td>
<td>0.00</td>
</tr>
<tr>
<td>Overall (I-squared = 44.3%, p = 0.084)</td>
<td>0.02 (-0.18, 0.21)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

NOTE: Weights are from random effects analysis
Figure 1

PRISMA flow diagram

5317 references identified through search
2390 duplicates removed

2927 studies screened against title and abstract
2822 studies excluded

105 studies assessed for full-text eligibility
71 studies excluded

34 studies included

41 did not measure facial emotion recognition
14 did not include those at high risk of schizophrenia based on family history or genetics
7 review papers
4 did not include a control group
3 Dissertations
1 conference abstract
1 genetic linkage study