# Supplementary Material

## Sub-Group Effect Analyses

The negative symptom measures all demonstrated a small but significant relationship with depressive symptoms (all measures included) in the sub-group analyses; the CAINS (k=5, pooled ES=0.188, 95% CI=0.046-0.330, z=2.59, p=0.01), SANS (k=18, pooled ES=0.281, 95% CI=0.219-0.343, z=8.87, p<.001) and PANSS-Neg (k=34, pooled ES=0.162, 95% CI=0.089-0.234, z=24.35, p<.001). The SANS showed the strongest relationship with depressive symptoms and the PANSS-Neg had the smallest effect size – this may reflect the larger number of items and therefore wider range of symptoms assessed by the SANS.

The depression measures also demonstrated a consistent small but significant relationship with negative symptoms (all measures included). The MADRS (k=5, pooled ES=0.334, 95% CI=0.162-0.506, z=3.81, p<.001), the BDI (k=9, pooled ES=0.198, 95% CI=0.135-0.261, z=26.14, p<.001), the HDRS (k=16, pooled ES=0.284, 95% CI=0.212-0.357, z=7.68, p<.001) and the CDSS (k=35, pooled ES=0.156, 95% CI=0.095-0.217, z=5.01, p<.001). the MADRS showed the strongest association with negative symptoms although the number of studies is small. The CDSS had the smallest effect size and this supports the conclusion of the Lako et al, 2012 review that this measure of depression most reliably distinguishes from negative symptoms in psychosis.

## Heterogeneity in Sub-Group Analyses

The negative symptom measures vary in the heterogeneity in the studies included, perhaps to some extent due to the different numbers of studies included in each group. The SANS shows low heterogeneity (p=.080, I2=34.5%, τ2=0.0055) and the CAINS is moderate (p=.046, I2=58.6%, τ2=0.0149). The PANSS-Neg, which is most commonly used has high heterogeneity at a similar level to the main effect analysis (p<.000, I2=81.6%, τ2=0.0308).

The depression measures showed moderate heterogeneity (ps<.05, I2=48.3-75.4%, τ2=0.0097-0.0242) with the exception of the BDI which showed very low heterogeneity amongst the 9 studies included in the analysis (p=.701, I2=0.0%, τ2=0.0000). These statistics represent 0% of the effect size attributed to between-study variance in studies using the BDI.

## Quality Assessment Tool for Quantitative Studies- Adapted Version

COMPONENT SCORES

A) SELECTION BIAS

(Q1) Are the individuals selected to participate in the study likely to be representative of the target population?

4 Very likely

3 Somewhat likely

2 Not likely

1 Can’t Tell

(Q2) What percentage of selected individuals agreed to participate?

4 80-100% agreement

3 60-79% agreement

1. <60% agreement

1 Can’t Tell

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Score this section | Very Strong  | Strong | Moderate | Weak |
|  | 4 | 3 | 2 | 1 |

B) STUDY DESIGN

(Q3) Was there a clear hypothesis stated and matching design for the study?

0 No

1 Yes

C) DATA COLLECTION METHODS

(Q4) Were data collection tools shown to be valid?

 3 Yes

2 No

1 Can’t tell

(Q5) Were data collection tools shown to be reliable?

3 Yes

2 No

1 Can’t tell

(Q6) Are dimensional scales reported for negative symptoms?

1. No
2. Yes

|  |  |  |  |
| --- | --- | --- | --- |
| Score this section | Strong | Moderate | Weak |
|  | 3 | 2 | 1 |

D) MISSING DATA

(Q7) Were missing data reported in terms of numbers and/or reasons?

3 Yes

2 No

1 Can’t tell

(Q8) Indicate the percentage of participants completing the study.

 4 80 -100%

3 60 - 79%

2 less than 60%

1 Can’t tell

|  |  |  |  |
| --- | --- | --- | --- |
| Score this section | Strong | Moderate | Weak |
|  | 3 | 2 | 1 |

E) ANALYSES

(Q9) Are the statistical methods appropriate for the study design?

3 Yes

1. No

1 Can’t tell

(Q10) Was the significance level adjusted appropriately for the number of comparisons being conducted?

1. Yes

2 No

1 Can’t Tell

|  |  |  |  |
| --- | --- | --- | --- |
| Score this section | Strong | Moderate | Weak |
|  | 3 | 2 | 1 |

## Quality Assessment Tool for Quantitative Studies: Guidance for Scoring

**A) SELECTION BIAS**

(Q1) Participants are more likely to be representative of the target population if they are randomly selected from a comprehensive list of individuals in the target population (score very likely).

They may not be representative if they are referred from a single setting (e.g. outpatient or inpatients only) in a systematic manner (score somewhat likely).

The sample is even less likely to be representative if from a single ward or service or if they have self-referred (score not likely).

(Q2) Refers to the % of subjects in the control and intervention groups that agreed to participate in the study before they were assigned to intervention or control groups. If this is not reported then score this as “can’t tell”.

**A Overall Score(1-3)**

Very Strong: The selected individuals are very likely to be representative of the target population (Q1 is 4) and there is greater than 80% participation (Q2 is 4).

Strong: The selected individuals are at least somewhat likely to be representative of the target population (Q1 is 4 or 3); and there is 60 - 79% participation (Q2 is 3).

 Moderate: The selected individuals are at least somewhat likely to be representative of the target population (Q1 is 4 or 3); and participation is not described (Q2 is 1) or below <60% (Q2 is 2).

Weak: The selected individuals are not likely to be representative of the target population (Q1 is 2); or there is less than 60% participation (Q2 is 2) or selection is not described (Q1 is 1); and the level of participation is not described (Q2 is 1).

**B) STUDY DESIGN**

Q3) Important to consider whether a clear hypothesis was stated and the design of the study was appropriate to address this (Yes/No).

**C) DATA COLLECTION METHODS**

Q4,5) Tools for primary outcome measures must be described as reliable and valid. If ‘face’ validity or ‘content’ validity has been demonstrated, this is acceptable.

Reliability and validity can be reported in the study or in a separate study. For example, some standard assessment tools have known reliability and validity.

Additional question inserted regarding the method of reporting the subscales as dimensional scales considered to be of a higher standard particularly in the field of negative symptoms.

**C Overall Score (1-3)**

Strong: The data collection tools have been shown to be valid (Q4 is 3); and the data collection tools have been shown to be reliable (Q5 is 3). Dimensions are reported rather than only total scale scores.

Moderate: The data collection tools have been shown to be valid (Q4 is 3); and the data collection tools have not been shown to be reliable (Q5 is 2) or reliability is not described (Q5 is 1). Dimensions are either reported or omitted (1 or 0).

Weak: The data collection tools have not been shown to be valid (Q4 is 2) or both reliability and validity are not described (Q4 is 1 and Q5 is 1).

**D) WITHDRAWALS AND DROP-OUTS**

Q 7,8) Score YES if the authors describe BOTH the numbers and reasons for withdrawals and drop-outs. Score NO if either the numbers or reasons for withdrawals and drop-outs are not reported.

The percentage of participants completing the study refers to the % of subjects remaining in the study at the final data collection period in all groups (i.e. control and intervention groups).

**D Overall Score (1-3)**

Strong: will be assigned when the completion rate is 80% or greater (Q7 is 4).

Moderate: will be assigned when the completion rate is 60 – 79% (Q8 is 3).

Weak: will be assigned when a completion rate is less than 60% (Q8 is 2) or any missing data were not described (Q8 is 1).

**E) ANALYSIS APPROPRIATE TO QUESTION**

**E Overall Score (1-3)**

Strong: will be assigned when the analysis is appropriate and the significance level accounts for the number of comparisons conducted (Q9 and 10 are both 3)

Moderate: will be assigned when the analysis is appropriate but the significance level has not been adapted. (Q9 is 3 and Q10 is 2)

Weak: will be assigned the suitability of the analyses and the adjustment of the significance level is not clear (Q9 and 10 are 1).

**Maximum Possible Score: 14**

Original sections: C,D and G have been omitted as the only apply to studies which consider between-groups analyses. This is not relevant to the research question considered in this review.