Introduction

Theory of mind (ToM) or 'mentalising' refers to the cognitive ability to attribute mental states such as thoughts, beliefs and intentions to other people. ToM thus allows an individual to explain and predict behaviour and to construct and understand complex social situations. ToM ability may be deficient in people with schizophrenia.

Several tasks have been developed to test ToM abilities. The most commonly used tasks include; false belief and deception tasks which test one’s ability to understand that someone else can hold a belief that is different from actual reality and act upon that false belief. A first-order ToM task of this type requires participants to understand that someone else is acting on a false belief. A second-order ToM task involves participants having to infer what one person thinks about what another person thinks. Another type of task commonly used is an “intention-inferencing” task, in which a person is required to infer what another person intends in situations where such information is not directly available. These situations might be described verbally, in a short story, or depicted visually in a cartoon or a cartoon sequence.

Other ToM tasks test capacities to understand “indirect” speech, such as when someone uses hints, jokes, irony or sarcasm. In the hinting task, participants are asked to infer the real intentions behind people’s speech. In all such cases, the listener needs to make inferences beyond the literal meanings of the words to appreciate what the speaker really intends to say. Another ToM task, the ‘eyes task’ tests the ability to infer another person’s complex mental states (e.g. embarrassment) from pictures which just show the eye region of that other person’s face. This test depends on more automatic decoding abilities rather than reasoning about mental states, unlike the previous tasks.

Method

We have included only data from systematic reviews (systematic literature search, detailed methodology with inclusion/exclusion criteria) and published in full text, in English, from the year 2000, in which results are reported separately for people with diagnoses of schizophrenia, schizoaffective disorder, schizophreniform disorder or first episode schizophrenia. Due to the high volume of systematic reviews we have now limited inclusion to systematic meta-analyses. Where no systematic meta-analysis exists for a topic, systematic reviews without meta-analysis are included for that topic. As part of a wider search for all topics included in the library, reviews on ToM for schizophrenia were identified by searching the databases MEDLINE, EMBASE, CINAHL, Current Contents, PsycINFO and the Cochrane library. Hand searching reference lists of identified reviews was also conducted. When multiple copies of reviews were found, only the most recent version was included. The decision to include or exclude reviews was conducted in duplicate by two reviewers with any disagreements settled by discussion. All quality assessments and data extraction have been completed in duplicate by two independent reviewers who were not masked to review authors.

Review reporting assessment was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist (formerly the QUOROM statement) which describes a preferred way to present a meta-analysis. Reviews were assigned a low, medium or high possibility of reporting bias depending on how many items were checked. For instance, a low possibility of bias would be assigned to reviews checking over 66% of items, a medium possibility between 33 and 66% and a high possibility would be given to reviews checking less than 33%. Due to the increased number of reviews published since
2014, reviews reporting less than 50% of items have been excluded from the library, prior to this date we excluded reviews reporting less than 33% of items. The PRISMA flow diagram is a suggested way of providing information about studies included and excluded with reasons for exclusion. Where no flow diagram has been presented by individual reviews, but identified studies have been described in the text, reviews have been checked for this item. Note that early reviews may have been guided by less stringent reporting checklists than the PRISMA, and that some reviews may have been limited by journal guidelines.

Evidence was graded using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group approach where high quality evidence such as that gained from randomized controlled trials (RCT) may be downgraded to moderate, low or very low if review and study quality is limited, if there is inconsistency in results, indirect comparisons, imprecise or sparse data and high probability of reporting bias. It may also be downgraded if risks associated with the intervention or other matter under review are high. Conversely, low quality evidence such as that gained from observational studies may be upgraded if effect sizes are large, there is a dose dependent response or if results are reasonably consistent, precise and direct with low associated risks (see end of table for an explanation of these terms). The resulting table represents an objective summary of the available evidence, although the conclusions are solely the opinion of staff of the Schizophrenia Research Institute.

Conclusions

- Moderate quality evidence suggests a large effect of impairment on ToM tasks in patients with schizophrenia compared to controls. Longer duration of illness and lower general intelligence reduce performance.
- Moderate to high quality evidence suggests a large effect of impairment on ToM tasks in people with first-episode psychosis, a medium effect in people at ultra-high risk of psychosis, and a small effect in relatives of patients with schizophrenia compared to controls.
- High quality evidence suggests a strong relationship between better performance on ToM tasks and better community functioning. Moderate to low quality evidence suggests better performance is also related to increased social skills, but not social behaviour.
- Moderate quality evidence suggests worse performance on ToM tasks may be strongly associated with increased disorganized and negative symptoms, but not associated with positive symptoms.
- Moderate quality evidence suggests that olanzapine and clozapine may improve patients’ performance on ToM tasks.

Results

We found seven systematic reviews that met our inclusion criteria. See PRISMA checklists for assessment of reporting transparency.
TECHNICAL COMMENTARY

Signs and Symptoms - Theory of Mind

_Bora E, Yucel M, Pantelis C_  
**Theory of mind impairment in schizophrenia: Meta-analysis**

[View review abstract online](#)

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Assessing performance on Theory of Mind (ToM) tasks in patients with a diagnosis of schizophrenia spectrum disorder versus healthy controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary of evidence</td>
<td>Moderate quality evidence (large samples, precise, inconsistent, possible publication bias) suggests a ToM impairment in people with schizophrenia across multiple tasks regardless of sex or age. Patients in an acute phase of the illness (inpatients) perform worse than outpatients or patients soon to be discharged from hospital. Lower general IQ contributed to lower ToM scores only in the patients in the remission phase</td>
</tr>
</tbody>
</table>

**Combined ToM score**

†Large effect showing impaired performance on combined ToM tasks score  
36 studies, N = 2117, $d = 1.10$, $95\% CI = 0.95$ to $1.25$, $p < 0.0001$, $p$ for Q test $< 0.001$  
$p$ for Eggers test for asymmetrical funnel plot $= 0.02$

Subgroup analysis investigating the effects of acute symptomatology; remitted (outpatients or inpatients just before discharge) versus non-remitted (inpatients) on combined ToM score

Significant difference between the two groups; $p < 0.0001$, $Q = 68.9$, $p < 0.0001$, $I^2 = 49.2\%$

Remitted patients; 16 studies, N = 984, $d = 0.80$, $95\% CI = 0.57$ to $1.03$, $p < 0.0001$, $p$ for Q test $< 0.004$  
$p$ for Eggers test for asymmetrical funnel plot $= 0.67$

Non-remitted patients; N = not reported  
$d = 1.21$, $95\% CI = 1.05$ to $1.37$, $p < 0.0001$, $p$ for Q test, $p$ for Eggers test not reported

Meta-regression investigating possible confounding effects of sex (33 studies), age (34 studies), education (17 studies), duration of illness (28 studies), antipsychotic dose (10 studies) and general intelligence (28 studies) on combined ToM score

No impact of sex and age on the results. Patients with longer duration of illness ($B = 0.03$, $p = 0.06$) and lower general intelligence ($B = 0.035$, $p = 0.06$) tended to be more impaired. The relationship
between general intelligence and ToM tasks became significant when the analysis was restricted to remitted patients ($B = 0.071$, $p = 0.01$)

<table>
<thead>
<tr>
<th>Task</th>
<th>Effect Size</th>
<th>CI</th>
<th>p-value</th>
<th>Asymmetry Test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hinting task</strong></td>
<td>Large</td>
<td>$d = 1.06$, $95% CI = 0.78$ to $1.34$, $p &lt; 0.0001$, $p$ for Q test &lt; 0.04</td>
<td>$p$ for Eggers test for asymmetrical funnel plot = 0.04</td>
<td></td>
</tr>
<tr>
<td><strong>Eyes task</strong></td>
<td>Large</td>
<td>$d = 0.90$, $95% CI = 0.64$ to $1.17$, $p &lt; 0.0001$, $p$ for Q test &lt; 0.194</td>
<td>$p$ for Eggers test for asymmetrical funnel plot = 0.15</td>
<td></td>
</tr>
<tr>
<td><strong>False Belief – Sequencing task</strong></td>
<td>Large</td>
<td>$d = 1.08$, $95% CI = 0.72$ to $1.43$, $p &lt; 0.0001$, $p$ for Q test &lt; 0.001</td>
<td>$p$ for Eggers test for asymmetrical funnel plot = 0.04</td>
<td></td>
</tr>
<tr>
<td><strong>False Belief – Stories task</strong></td>
<td>Large</td>
<td>$d = 1.06$, $95% CI = 0.76$ to $1.37$, $p &lt; 0.0001$, $p$ for Q test &lt; 0.001</td>
<td>$p$ for Eggers test for asymmetrical funnel plot = 0.06</td>
<td></td>
</tr>
</tbody>
</table>

Consistency in results: Inconsistent

Precision in results: Precise

Directness of results: Direct

---

**Bora E, Pantelis C**

*Theory of mind impairments in first-episode psychosis, individuals at ultra-high risk for psychosis and in first-degree relatives of schizophrenia: Systematic review and meta-analysis*
Comparison 1

Assessing performance on Theory of Mind tasks in relatives of patients with a diagnosis of schizophrenia spectrum disorder versus healthy controls

Summary of evidence

Moderate to high quality evidence (large samples, precise, consistent, direct) suggests a small to medium sized ToM impairment in relatives of people with schizophrenia across verbal and visual tasks. Lower education contributed to lower ToM scores.

Combined ToM score

Small to medium effect showing impaired performance in relatives vs. controls

12 studies, N = 3117, $d = 0.37$, 95%CI 0.19 to 0.54, $p < 0.001$, $I^2$ 0%, Fail-safe N = 146

Note: Meta-regression analyses suggested that longer duration of education in the control groups vs. relatives explains some of the between-group differences ($B = 0.41$, $p = 0.002$)

ToM verbal

Small effect showing impaired performance in relatives vs. controls

8 studies, N = 2946, $d = 0.24$, 95%CI 0.13 to 0.33, $p < 0.001$, $I^2$ 0%

ToM visual

Small to medium effect showing impaired performance in relatives vs. controls

9 studies, N = 555, $d = 0.36$, 95%CI 0.10 to 0.63, $p < 0.001$, $I^2$ 0%

Eyes task

No differences between groups

5 studies, N = 261, $d = 0.19$, 95%CI -0.10 to 0.48, $p = 0.19$, $I^2$ 0%

Consistency in results

Consistent

Precision in results

Precise

Directness of results

Direct
## Signs and Symptoms - Theory of Mind

<table>
<thead>
<tr>
<th>Comparison 2</th>
<th>Assessing performance on Theory of Mind (ToM) tasks in people at ultra-high risk of psychosis versus healthy controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary of evidence</td>
<td>Moderate to high quality evidence (large samples, precise, consistent, direct) suggests a medium sized ToM impairment in people at ultra-high risk of psychosis across verbal and visual tasks</td>
</tr>
</tbody>
</table>

### Combined ToM score

*Medium effect showing impaired performance in ultra-high risk groups vs. controls*

- 7 studies, $N = 581$, $d = 0.45$, 95%CI 0.23 to 0.67, $p < 0.001$, $I^2$ 0%, Fail-safe N = 42

#### ToM verbal

*Medium effect showing impaired performance in ultra-high risk groups vs. controls*

- 4 studies, $N = 329$, $d = 0.49$, 95%CI 0.26 to 0.72, $p < 0.001$, $I^2$ 0%

#### ToM visual - eyes

*Medium effect showing impaired performance in ultra-high risk groups vs. controls*

- 6 studies, $N = 497$, $d = 0.40$, 95%CI 0.14 to 0.70, $p = 0.003$, $I^2$ 0%

### Consistency in results

Consistent

### Precision in results

Precise

### Directness of results

Direct

<table>
<thead>
<tr>
<th>Comparison 3</th>
<th>Assessing performance on Theory of Mind (ToM) tasks in people with first-episode psychosis versus healthy controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary of evidence</td>
<td>Moderate to high quality evidence (large samples, precise, consistent, direct) suggests a large sized ToM impairment in people with first-episode psychosis across verbal and visual tasks</td>
</tr>
</tbody>
</table>

### Combined ToM score

*Large effect showing impaired performance in first-episode psychosis groups vs. controls*

- 8 studies, $N = 513$, $d = 1.00$, 95%CI 0.81 to 1.18, $p < 0.001$, $I^2$ 0%, Fail-safe N = 214

#### ToM verbal
Signs and Symptoms - Theory of Mind

<table>
<thead>
<tr>
<th>Large effect showing impaired performance in first-episode psychosis groups vs. controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 studies, N = 320, d = 0.99, 95%CI 0.76 to 1.23, p &lt; 0.001, I² 0%</td>
</tr>
</tbody>
</table>

**ToM visual - eyes**

<table>
<thead>
<tr>
<th>Large effect showing impaired performance in first-episode psychosis groups vs. controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 studies, N = 261, d = 0.94, 95%CI 0.69 to 1.20, p &lt; 0.001, I² 0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Consistency in results</th>
<th>Consistent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision in results</td>
<td>Precise</td>
</tr>
<tr>
<td>Directness of results</td>
<td>Direct</td>
</tr>
</tbody>
</table>


_The Functional Significance of Social Cognition in Schizophrenia: A Review_


View review abstract online

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Effect of deficits in Theory of Mind on functional outcome in schizophrenia spectrum disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary of evidence</td>
<td>Moderate to low quality evidence (direct, large sample size, unable to assess consistency, precision) suggests that theory of mind ability only influenced social skills, but not social behaviour or community function. Premorbid social function had some bearing on theory of mind ability.</td>
</tr>
</tbody>
</table>

Theory of Mind (ToM) and functional outcome
Signs and Symptoms - Theory of Mind

Theory of Mind is the cognitive ability to attribute mental states such as thoughts, beliefs and intentions to other people.

One study (N = 23) found limited evidence for a small significant association between ToM and social behaviour.

One study (N = 49) found ToM had a significant medium size association with overall social skill in outpatients.

One study (N = 44) found evidence for a medium significant association between increased ToM and community function.

One study (N = 42) found a large significant relationship between ToM and premorbid social functioning.

<table>
<thead>
<tr>
<th>Consistency in results</th>
<th>Unable to assess consistency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision in results</td>
<td>Unable to assess precision</td>
</tr>
<tr>
<td>Directness of results</td>
<td>Direct</td>
</tr>
</tbody>
</table>

Fett, A-K., Viechtbauer, W., Dominguez, M., Penn, D., van Os, J. and Krabbendam, L.

The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: A meta-analysis

Neuroscience and Biobehavioural Reviews, 2011. 35: 573-588

View review abstract online

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Association between functional outcome and cognitive performance on Theory of Mind (ToM) tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary of evidence</td>
<td>Moderate to high quality evidence (direct, consistent, precise) suggests a medium association between increased performance on ToM tasks and better community functioning in patients diagnosed with schizophrenia</td>
</tr>
</tbody>
</table>

Community functioning

- 3 studies, N = 114
- Significant medium to large association between increased performance on ToM tasks and better community functioning
Estimated average correlation = 0.48, 95%CI 0.32 to 0.61, \( p < 0.001 \)

\[ Q = 0.81, I^2 = 1\%, \text{ non-significant (p value not reported)} \]

<table>
<thead>
<tr>
<th>Consistency in results</th>
<th>Consistent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision in results</td>
<td>Precise</td>
</tr>
<tr>
<td>Directness of results</td>
<td>Direct</td>
</tr>
</tbody>
</table>

**Pickup G.J.**

**Relationship between Theory of Mind and executive functioning in schizophrenia: A systematic review**

*Psychopathology, 2008. 41: 206-213*

[View review abstract online](#)

<table>
<thead>
<tr>
<th>Comparison</th>
<th>The association between ToM and executive functioning in people with schizophrenia vs. controls (healthy and various control groups)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary of evidence</td>
<td>Moderate to low quality evidence (unable to assess consistency or precision) suggests impaired performance on various Theory of Mind tasks in people with schizophrenia compared to controls, which may be associated with greater symptom severity. The evidence also suggests that Theory of Mind tasks may be correlated with executive functioning tasks</td>
</tr>
</tbody>
</table>

**Performance on Theory of Mind tasks**

Authors reported poorer performance in people with schizophrenia compared to controls on:

- False Beliefs (FB) picture sequencing tasks (8 studies, \( N = 392 \)), ToM questionnaire and vignettes (3 studies, \( N = 201 \)), comprehension (1 study, \( N = 54 \)), hinting tasks (2 studies, \( N = 220 \)), 1st- and 2nd-order FB stories (4 studies, \( N = 201 \)), and 1st-order FB stories (1 study, \( N = 60 \)).

**ToM association with executive functioning**

Authors reported that six studies (\( N = 227 \)) suggest overall ToM performance was associated with executive functioning tasks, including capture picture-sequencing (\( N = 56 \)), Weigl (\( N = 60 \)), key search and zoo map (\( N = 41 \)) and trails B (\( N = 50 \)). 4 studies reported no association with ToL (\( N = \))
Signs and Symptoms - Theory of Mind

3 studies (N = 148) suggest FB scores were associated with executive functioning tasks, including capture (2 studies, N = 90), ToL (1 study, N = 45) and WCST (1 study, N = 58).

1 study (N = 128) reported that hinting scores correlated negatively with Trails B and positively with the WCST task.

**ToM association with symptoms**

Poorer ToM performance was associated with greater severity of negative symptoms (2 studies, N = 107), positive symptoms (1 study, N = 40), paranoid symptoms (1 study, N = 63), poor insight (1 study, N = 55), behavioural problems (1 study, N = 60), disorganised symptoms (2 studies, N = 94), psychomotor poverty (1 study, N = 52) and positive formal thought disorder (3 studies, N = 236).

Better ToM performance was associated with better social functioning (3 studies, N = 201).

However, two studies reported that ToM was not associated with general symptoms (N = 127), negative symptoms (N = 56) or paranoid symptoms (1 study, N = 56).

Increased negative and positive symptom severity was associated with poorer performance on the hinting task (1 study, N = 50), and increased negative symptoms were associated with poorer picture sequencing and capture errors performance (1 study, N = 56).

Greater social functioning was associated with better performance on the reading the mind in the eyes task (1 study, N = 50), and better 2nd order FB scores were associated with greater insight (N = 1 study, N = 58).

<table>
<thead>
<tr>
<th>Consistency in results</th>
<th>Unable to assess consistency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision in results</td>
<td>Unable to assess precision</td>
</tr>
<tr>
<td>Directness of results</td>
<td>Direct</td>
</tr>
</tbody>
</table>

So, S., Garety, P., Peters, E. & Kapur, S.

**Do antipsychotics improve reasoning bias? A review**

*Psychosomatic Medicine* 2010. 72:681-693

[View review abstract online](#)

**Comparison**

Relationship between performance on Theory of Mind (ToM) tasks, symptoms and antipsychotic medication type

**Summary of evidence**

Moderate quality evidence (direct, unable to assess precision) suggests performance on ToM tasks may be associated with
negative and total scores on the PANSS and SANS, with positive symptoms showing a weaker association

Moderate to low quality evidence (direct, unable to assess precision) suggests performance on ToM stories may be related to delusional outcome (preoccupation and distress)

Moderate quality evidence (direct, unable to assess precision) suggests antipsychotics may improve performance on ToM tasks for patients on olanzapine and clozapine

<table>
<thead>
<tr>
<th>Positive and negative symptoms</th>
<th>Measured by PANSS and SANS</th>
</tr>
</thead>
</table>

1 observational study, N = 128 (medication not reported), measured performance on a hinting task

ToM performance correlated with PANSS positive, negative and delusion scores, with increased performance being related to reduction in symptoms. Patients with disorganized symptoms tended to perform worse than those with positive and negatives symptoms

1 observational study, N = 77 (22 patients on antipsychotics vs. 55 controls), measured performance on a computerised mental inference task

Affective ToM was associated with negative symptoms (SANS alogia, SANS attention and SANS total symptoms), whereas cognitive ToM was associated with positive symptoms (PANSS). The authors concluded that ToM performance was more strongly related to negative symptoms and thought disorder than to positive symptoms

1 observational study, N = 71 (on atypical (88.6%) and typical (11.4%) antipsychotics), measured performance on a hinting task.

ToM performance correlated with PANSS negative symptoms, general and total score, but not positive score on the PANSS

Delusions
Signs and Symptoms - Theory of Mind

1 observational study, N = 21 (10 fully remitted, 5 partially remitted, 6 acutely deluded), 15 patients were on antipsychotics, measured performance on a picture-sequencing task and ToM questionnaire

No group differences, authors concluded that ToM performance was not related to severity of delusions

1 observational study, N = 128 (39 currently paranoid, 29 remitted paranoid, 27 non-psychotic depressed, 33 healthy controls), measured performance on ToM stories

Patients with persecutory delusions scored lower on the ToM task. Performance on ToM stories correlated significantly with delusional preoccupation and distress, but the picture-sequencing task did not and there was no correlation with antipsychotic dosage

<table>
<thead>
<tr>
<th>Effects of antipsychotic medications on task performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 observational study, N = 108 (84 patients (77 schizophrenia, 7 schizoaffective) and 24 healthy controls), on first generation/clozapine/olanzapine/risperidone for 4 months, most on mood stabilizers or other medications and performance on 1st- and 2nd- order belief task and faux-pas task</td>
</tr>
<tr>
<td>Mean score on BPRS was highest in the clozapine group, and lowest in those receiving first generation antipsychotics and olanzapine</td>
</tr>
<tr>
<td>The olanzapine and clozapine groups performed similarly to controls on the ToM task, but those on first generation antipsychotics and risperidone performed worse than the other groups</td>
</tr>
</tbody>
</table>

1 longitudinal study, N = 17 patients who were drug-free then started on clozapine/olanzapine/risperidone/ioxapine, measured performance on a hinting task

ToM performance improved with antipsychotic use, particularly during the first 2 weeks of treatment

No relationship between change in ToM and change in symptoms

Consistency in results | N/A – no meta-analysis |
---|---
Precision in results | No measure of precision reported |
Directness of results | Direct |

Sprong M, Schothorst P, Vos E, Hox J, van Engeland H

Theory of mind in schizophrenia. Meta-analysis

### Signs and Symptoms - Theory of Mind

**Comparison 1**

<table>
<thead>
<tr>
<th>Assessing performance on Theory of Mind (ToM) tasks in patients with a diagnosis of schizophrenia spectrum disorder versus healthy controls</th>
</tr>
</thead>
</table>

**Summary of evidence**

<table>
<thead>
<tr>
<th>Moderate to high quality evidence (large samples, consistent, precise), show impairment on ToM tasks regardless of IQ, sex or age compared to healthy controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients with disorganized symptoms are particularly impaired on these tasks</td>
</tr>
</tbody>
</table>

**Combined ToM score**

<table>
<thead>
<tr>
<th>All participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large effect showing impaired performance on combined ToM tasks score compared to controls</td>
</tr>
<tr>
<td>29 observational studies, N = 831</td>
</tr>
<tr>
<td>$d = -1.25, 95% CI = -1.441$ to $-1.069, \quad p &lt; 0.0001, \quad Q = 29.13, \quad p$ for Q test $&lt; 0.41, \quad I^2 = 3%$</td>
</tr>
<tr>
<td>Regression analysis showed no difference in results when IQ, sex and age added to analysis</td>
</tr>
<tr>
<td>Note; wherever patients performed poorer than controls, effect sizes are negative</td>
</tr>
<tr>
<td>Fail-safe number = 153, indicating that 153 non-significant, unpublished studies are required to reduce the effect to a negligible level (no publication bias)</td>
</tr>
</tbody>
</table>

**Subgroup analysis; disorganised symptoms**

| Large effect showing impaired performance on combined ToM tasks score compared to controls |
| 9 observational studies, N not reported |
| $d = -2.231, 95\% CI = -2.565$ to $-1.897, \quad p, \quad Q$ not reported, authors state data homogenous |
| Patients with disorganised symptoms were also significantly more impaired than those without disorganized symptoms, those with paranoid symptoms and those in remission |

**Subgroup analysis; patients in remission**

| Medium effect showing impaired performance on combined ToM tasks score compared to controls |
| 5 observational studies, N not reported |
| $d = -0.693, 95\% CI = -1.017$ to $-0.367, \quad p < 0.01, \quad Q = 7.3816, \quad p < 0.05, \quad I^2 = 45.8\%$ |

**Consistency in results**

| Consistent, except for patients in remission |

**Precision in results**

| All results precise |

**Directness of results**

| Direct for people with schizophrenia spectrum diagnosis compared to healthy controls |
Signs and Symptoms - Theory of Mind

Explanation of acronyms

B = coefficient, CI = Confidence Interval, $d = \text{Cohen's} \ d$ and $g = \text{Hedges'} \ g = \text{standardized mean differences (see below for interpretation of effect size)}$, Eggers test = test for asymmetry for detecting possible publication bias, FB = False Beliefs, $I^2 = \text{the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance)}$, N = number of participants, $p = \text{statistical probability of obtaining that result (} p < 0.05 \text{ generally regarded as significant)}$, PANSS = Positive and Negative Syndrome Scale, Q = Q statistic (chi-square) for the test of heterogeneity in results across studies, SANS = Scale of the Assessment of Negative Symptoms, SCWT = Stroop colour word test, ToL = Tower of London, ToM = Theory of Mind, WCST = Wisconsin Card Sorting Task
Explanation of technical terms

* Bias has the potential to affect reviews of both RCT and observational studies. Forms of bias include; reporting bias – selective reporting of results, publication bias - trials which are not formally published tend to show less effect than published trials, further if there are statistically significant differences between groups in a trial, these trial results tend to get published before those of trials without significant differences; language bias – only including English language reports; funding bias - source of funding for the primary research with selective reporting of results within primary studies; outcome variable selection bias; database bias - including reports from some databases and not others; citation bias - preferential citation of authors. Trials can also be subject to bias when evaluators are not blind to treatment condition and selection bias of participants if trial samples are small11.

† Different effect measures are reported by different reviews.

Prevalence refers to how many existing cases there are at a particular point in time. Incidence refers to how many new cases there are per population in a specified time period. Incidence is usually reported as the number of new cases per 100,000 people per year. Alternatively some studies present the number of new cases that have accumulated over several years against a person-years denominator. This denominator is the sum of individual units of time that the persons in the population are at risk of becoming a case. It takes into account the size of the underlying population sample and its age structure over the duration of observation.

Reliability and validity refers to how accurate the instrument is. Sensitivity is the proportion of actual positives which are correctly identified (100% sensitivity = correct identification of all actual positives) and specificity is the proportion of negatives which are correctly identified (100% specificity = not identifying anyone as positive if they are truly not).

Weighted mean difference scores refer to mean differences between treatment and comparison groups after treatment (or occasionally pre to post treatment) and in a randomized trial there is an assumption that both groups are comparable on this measure prior to treatment. Standardized mean differences are divided by the pooled standard deviation (or the standard deviation of one group when groups are homogenous) which allows results from different scales to be combined and compared. Each study’s mean difference is then given a weighting depending on the size of the sample and the variability in the data. Less than 0.4 represents a small effect, around 0.5 a medium effect, and over 0.8 represents a large effect11.

Odds ratio (OR) or relative risk (RR) refers to the probability of a reduction (< 1) or an increase (> 1) in a particular outcome in a treatment group, or a group exposed to a risk factor, relative to the comparison group. For example, a RR of 0.75 translates to a reduction in risk of an outcome of 25% relative to those not receiving the treatment or not exposed to the risk factor. Conversely, a RR of 1.25 translates to an increased risk of 25% relative to those not receiving treatment or not having been exposed to a risk factor. A RR or OR of 1.00 means there is no difference between groups. A medium to large effect is considered if RR > 2 or < 0.5 and a large effect if RR > 5 or < 0.212. lnOR stands for logarithmic OR where a lnOR of 0
shows no difference between groups. Hazard ratios measure the effect of an explanatory variable on the hazard or risk of an event.

Correlation coefficients (e.g., r) indicate the strength of association or relationship between variables. They can provide an indirect indication of prediction, but do not confirm causality due to possible and often unforseen confounding variables. An r of 0.10 represents a weak association, 0.25 a medium association and 0.40 and over represents a strong association.

Unstandardized (b) regression coefficients indicate the average change in the dependent variable associated with a 1 unit change in the independent variable, statistically controlling for the other independent variables. Standardized regression coefficients represent the change being in units of standard deviations to allow comparison across different scales.

‡ Inconsistency refers to differing estimates of treatment effect across studies (i.e. heterogeneity or variability in results) which is not explained by subgroup analyses and therefore reduces confidence in the effect estimate. $I^2$ is the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance) - 0% to 40%: heterogeneity might not be important, 30% to 60%: may represent moderate heterogeneity, 50% to 90%: may represent substantial heterogeneity and 75% to 100%: considerable heterogeneity. $I^2$ can be calculated from Q (chi-square) for the test of heterogeneity with the following formula:\cite{13};

$$I^2 = \left( \frac{Q - df}{Q} \right) \times 100\%$$

$§$ Imprecision refers to wide confidence intervals indicating a lack of confidence in the effect estimate. Based on GRADE recommendations, a result for continuous data (standardised mean differences, not weighted mean differences) is considered imprecise if the upper or lower confidence limit crosses an effect size of 0.5 in either direction, and for binary and correlation data, an effect size of 0.25. GRADE also recommends downgrading the evidence when sample size is smaller than 300 (for binary data) and 400 (for continuous data), although for some topics, this criteria should be relaxed\cite{12}.

¶ Indirectness of comparison occurs when a comparison of intervention A versus B is not available but A was compared with C and B was compared with C which allows indirect comparisons of the magnitude of effect of A versus B. Indirectness of population, comparator and or outcome can also occur when the available evidence regarding a particular population, intervention, comparator, or outcome is not available so is inferred from available evidence. These inferred treatment effect sizes are of lower quality than those gained from head-to-head comparisons of A and B.

Many thanks to Associate Professor Robyn Langdon for reviewing the original version of this summary of evidence

Macquarie Centre for Cognitive Science, Macquarie University, Sydney
TECHNICAL
COMMENTARY

Signs and Symptoms - Theory of Mind

References