Chapter 6

Metacognitive training for schizophrenia spectrum patients: a meta-analysis on outcome studies

B.J. van Oosterhout
F. Smit
A.C. Krabbendam
S. Castelein
A.B.P. Staring
M. van der Gaag


B.J. van Oosterhout was involved in developing the study concept and design; in the management of the study; the acquisition, analysis, and interpretation of the data; and in drafting and revising the manuscript.
Abstract

BACKGROUND: Metacognitive training (MCT) for schizophrenia spectrum is widely implemented. It is timely to systematically review the literature and to conduct a meta-analysis.

METHODS: Eligible studies were selected from several sources (databases and expert suggestions). Criteria included comparative studies with a MCT-condition measuring positive symptoms and/or delusions and/or data gathering bias. Three meta-analyses were conducted on data gathering (3 studies; 219 participants), delusions (7 studies; 500 participants) and positive symptoms (9 studies; 436 participants). Hedge’s g is reported as the effect size of interest. Statistical power was sufficient to detect small-to-moderate effects.

RESULTS: All analyses yielded small non-significant effect sizes (0.26 for positive symptoms; 0.22 for delusions; 0.31 for data gathering bias). Corrections for publication bias further reduced the effect sizes to 0.21 for positive symptoms and to 0.03 for delusions. In blinded studies, the corrected effect sizes were 0.22 for positive symptoms and 0.03 for delusions. In studies using proper intention-to-treat statistics the effect sizes were 0.10 for positive symptoms and -0.02 for delusions. The moderate-to-high heterogeneity in most analyses suggests that processes other than MCT alone have an impact on the results.

CONCLUSIONS: The studies so far do not support a positive effect for MCT on positive symptoms, delusions and data-gathering. The methodology of most studies was poor and sensitivity analyses to control for methodological flaws reduced the effect-sizes considerably. More rigorous research would be helpful in order to create enough statistical power to detect small effect sizes and to reduce heterogeneity. Limitations and strengths are discussed.
**Introduction**

**Recent developments in CBT in psychosis**
Cognitive behavioral therapy (CBT) is recommended for psychosis in many national guidelines (e.g. NICE, 2009). Meta-analyses have demonstrated modest but robust positive effects in blinded CBT-trials on psychotic symptoms with small-to-moderate effect sizes (Zimmermann et al., 2005: Effect-size positive symptoms = 0.29; Wykes et al., 2008: Effect-size positive symptoms = 0.23; Burns et al., 2014: effect-size positive symptoms = 0.43; van der Gaag et al., 2014: Effect-size delusions = 0.24 and hallucinations = 0.46). One meta-analysis produced non-significant effect-size in blinded studies (Jauhar et al., 2014: Effect-size positive symptoms = 0.08). Moreover, CBT was superior to any other psychosocial intervention in reducing positive symptoms (Turner et al., 2014) and yielded robust results in all sensitivity analyses for risk of bias. It only disappeared in allegiance sensitivity analysis because of lack of power as only three studies were non-alleged. The focus recently moved to examination of the working mechanisms and cognitive biases. The biases associated with data gathering and the appraisal and processing of information, are associated to psychosis in general and, in particular, to positive symptoms such as (persecutory) delusions (van der Gaag, 2006; Freeman, 2007). The focus of therapies and trainings have moved from a predominantly content-oriented focus (what is the patient thinking?) towards a process-oriented focus on cognitive biases (Garety et al., 2001; Morrison, 2001; Bentall et al., 2009; Bennett and Corcoran, 2010).

**Cognitive biases associated with delusions**
Several cognitive biases, such as jumping-to-conclusions (JTC), belief inflexibility, problems in Theory of Mind and externalizing attributions, are hypothesized to be associated with the pathogenesis and maintenance of delusions. The JTC bias refers to a tendency to gather less data or evidence than healthy controls in order to reach a decision or accept a hypothesis (Garety et al., 1991; Fine et al., 2007). The JTC bias has also been found in individuals at risk for psychosis (Colbert and Peters, 2002; Van Dael et al., 2006) and in highly deluded and remitted patients (Moritz & Woodward, 2005). Belief inflexibility refers to a bias against disconfirmatory evidence (Woodward et al., 2008) and is particularly related to delusional preoccupation and conviction (Garety et al., 2005; Colbert et al., 2010). Furthermore, problems in Theory of Mind (i.e. the inability to represent the beliefs, thoughts and intentions of others), which is known to be related to symptoms of disorganization, may also contribute to paranoid delusions (Craig et al., 2004; Versmissen et al., 2008; Abdel-Hamid et al., 2009); however, there is mixed evidence on this topic with some studies finding no associations (Fernyhough et al., 2008) or an intact Theory of Mind during a delusional state (Walston et al., 2000). Finally, there is
evidence that an externalizing attribution style, with patients making external (personal) attributions for negative events and internal attributions for positive events, is associated with delusions (Kaney & Bentall, 1989; Kinderman & Bentall, 1997; Janssen et al., 2006). Again, inconsistent findings have been reported, with some studies finding no differences between early psychosis patients and controls in the tendency to externalize or personalize (Langdon et al., 2013) and others concluding that the link between persecutory ideation and attribution biases only manifests when persecutory ideation is of delusional intensity, and that it is confined to only a personalizing bias (McKay et al., 2007). Generally speaking, the above-mentioned cognitive biases are of interest since they are assumed to mediate (or moderate) treatment response in delusional symptoms (So et al., 2010).

**Metacognitive training in schizophrenia**

Moritz and Woodward were the first to translate theoretical results on cognitive biases and processes into a series of training modules called the Metacognitive Training (MCT; Moritz & Woodward (2007)). Furthermore, sessions on overconfidence in memory errors and depressive cognitive patterns were added. MCT aims to increase the patient’s knowledge about cognitive biases and to raise (metacognitive) awareness of the dysfunctional nature of these biases by means of exercises. It adopts a ‘back-door approach’ by first addressing cognitive biases instead of directly aiming at core delusional beliefs. MCT is a group-wise training for 3–10 patients and comprised of eight different modules targeting cognitive biases. Exercises that demonstrate the fallibility of human cognitive apparatus are discussed in the group. Participants are encouraged to express personal examples of these biases, and discussion of ways to counter them, serve to provide corrective experiences in a supportive atmosphere. This approach has obvious advantages over mere didactic providing of information. Patients are taught to recognize and confront the biases that are important in schizophrenia, thus allowing them to arrive at more appropriate inferences.

The published results on MCT are inconsistent and the evidence for efficacy is still undecided. At the same time there is a widespread dissemination and the MCT modules are available in 33 languages and are used all over the globe. Although the number of studies is relatively small for properly powered meta-analysis (N = 11), we considered it is necessary to systematically review the current literature and to conduct a meta-analysis on the effects of MCT compared to TAU or active control on data gathering bias, delusions, and positive symptoms of psychosis in patients with positive symptoms of schizophrenia.
Methods

Data collection

Eligibility criteria
Studies had to meet the following criteria for inclusion: a) the experimental treatment was MCT (Moritz/Woodward approach), b) the study had to be a comparative trial with or without randomization, c) reporting both pre and post test measures d) any control condition was accepted, e) at least 75% of the patients were diagnosed with schizophrenia spectrum disorders, f) only published in peer-reviewed journals (conference abstracts were excluded); and g) the study used data gathering, delusion ratings and/or positive symptom ratings as an outcome measure. Although there were no language restrictions all studies were in English language. The PRISMA guidelines for systematic reviews and meta-analyses are followed (Liberati et al., 2009).

Information sources
Studies were selected by various methods. First, a systematic search was made (from 2002 to 1st July 2014) in Medline, PsycINFO, EMBASE, the Cochrane Central Register of Controlled Trials and the Cochrane Database of Systematic Reviews, to detect studies describing metacognitive therapy or training in patients with positive symptoms of psychosis and schizophrenia.

Articles were identified by combining terms indicative of metacognitive psychological treatment (search terms included ‘metacognitive training’ OR ‘MCT’) AND outcome research (search terms included ‘randomised controlled trial’, ‘randomized controlled trial’ OR ‘RCT) (for the algorithm, please contact the first author).

Second, the search was supplemented by relevant papers identified by manual search of the reference lists of the identified articles. Finally, leading researchers in the field of CBT and MCT (Prof. S. Moritz, Prof. P. Garety, Prof. D. Freeman and Prof. A. Morrison) were asked to make suggestions regarding relevant literature.

Data extraction
The titles of the 611 retrieved papers were screened for eligibility by the first author. A first selection on the topic of ‘MCT in psychosis’ resulted in 22 potential papers.

Few studies were excluded based on the aforementioned criteria: Aghotor et al. (2010) (based on criterium c), Erawati et al. (2013) (based on criterium c), Favrod et al. (2011) (based on criterium b), Ferwerda et al. (2010) (based on criterium b), Moritz & Woodward (2007) (based on criterium g), Moritz et al. (2011b) (based on criterium g: no reports on total scores, only subscales).
Finally, 11 studies were included in this meta-analysis (Figure 6.1). One study had delusional symptoms as primary outcome (van Oosterhout et al., 2014), two had positive symptoms as primary outcome (Naughton et al., 2012; Balzan et al., 2014) and five papers dealt with both delusional symptoms and positive symptoms (Kumar et al., 2010; Moritz et al., 2011a, Favrod et al., 2014; Kuokkanen et al., 2014; Briki et al., 2014). One paper dealt with both data gathering bias and positive symptoms (Rocha & Queiros, 2013), one paper dealt with data gathering bias, positive symptoms and delusional symptoms (Moritz et al., 2013), and one paper had data gathering bias as primary outcome (Ross et al., 2011).
Table 6.1 lists the characteristics of the 11 selected studies. Studies differed in sample size from 16 (Kumar et al.) to 150 (Moritz et al.). Four trials had inpatients, four had outpatients and three had both. Training was either 8 or 16 sessions in most trials, but only short modules of only max. 60 minutes in two trials. Several researchers had made small adaptations to the MCT package. The trial by Moritz et al. (2011a) embedded the MCT within an individual CBT; Rocha & Queirós (2013) added training in social cognition; Balzan et al. (2014) had a single module focusing on data gathering and bias against disconfirmation, and Ross et al. (2011) had a cut-down single module version partially (2/3) based on MCT. So, there is quite some heterogeneity in the training format.

**Quality assessment**

The methodological quality of the eligible studies was reviewed using the Clinical Trial Assessment Measure (CTAM) (Tarrier & Wykes, 2004). Quality ratings were based on the following criteria: sample characteristics (i.e. is the sample a convenience sample or a geographically representative cohort; sample size); allocation procedures (i.e. valid randomization procedure); assessment of outcomes (i.e. standardized assessment method used); control condition (i.e. has a credible control condition been implemented); analysis (i.e. appropriate statistical analysis given the design and type of outcome); description of treatment (i.e. has the treatment been sufficiently described or manualized).

The maximum achievable score on the CTAM is 100. Similar to Wykes et al. (2008) we adopted an arbitrary cut-off score of 65 to denote either high or low quality studies. Two experienced independent raters (SC and ABPS) performed the screenings. A consensus meeting was held to resolve differences in scores and ratings.

**Data analysis**

Meta-analyses were conducted for the end-of-treatment effects for each of the available outcome measures separately. The outcomes at the end of treatment across the trials were synthesized meta-analytically using Comprehensive Meta-Analysis version 2.2 (www.meta-analysis.com). Post-hoc power analysis for random effect models in meta-analysis resulted in a detectable small-to-moderate pooled effect size of 0.37 (Hedge’s g; two-sided, power = 0.80, α = 0.05) in both positive symptoms and delusions.

The studies in this meta-analysis examined different samples and used various control interventions. Therefore, differences between the effect sizes are likely to reflect these sources of heterogeneity. A random effects model for meta-analytic synthesis of effect sizes across the primary studies was conducted. Most studies used small samples. We decided to use Hedges’ g as effect size, which is corrected for small sample bias.
### Table 6.1 Description of the interventions, patient characteristics, quality of the studies, and location

<table>
<thead>
<tr>
<th>Author, Year, Setting</th>
<th>Severity of symptoms delusions</th>
<th>Intervention</th>
<th>Number of sessions</th>
<th>N</th>
<th>Drop-out</th>
<th>Age mean (SD)</th>
<th>Male sex</th>
<th>Intervention</th>
<th>Number of sessions</th>
<th>N</th>
<th>Drop-out</th>
<th>Age mean (SD)</th>
<th>Male sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kumar et al. 2010</td>
<td>Acute symptoms</td>
<td>MCT</td>
<td>8 sessions</td>
<td>8</td>
<td>0%</td>
<td>31 (8.0)</td>
<td>n.a.</td>
<td>TAU</td>
<td>8</td>
<td>8</td>
<td>0%</td>
<td>34 (8.2)</td>
<td>n.a.</td>
</tr>
<tr>
<td>Moritz et al. 2011</td>
<td>Subacute symptoms</td>
<td>MCT + CBT</td>
<td>8 sessions</td>
<td>24</td>
<td>4%</td>
<td>33 (12.5)</td>
<td>71%</td>
<td>CogPack</td>
<td>8</td>
<td>24</td>
<td>12.5%</td>
<td>35 (9.1)</td>
<td>58%</td>
</tr>
<tr>
<td>Naughton et al. 2012</td>
<td>Mild-to-moderate</td>
<td>MCT</td>
<td>16 sessions</td>
<td>11</td>
<td>0%</td>
<td>37 (10.6)</td>
<td>100%</td>
<td>TAU</td>
<td>n.a.</td>
<td>8</td>
<td>0%</td>
<td>36 (1.2)</td>
<td>100%</td>
</tr>
<tr>
<td>Moritz et al. 2013</td>
<td>Mild to moderate ***</td>
<td>MCT</td>
<td>8 or 16 sessions</td>
<td>76</td>
<td>12%</td>
<td>37 (11.1)</td>
<td>59%</td>
<td>CogPack</td>
<td>Max 16 sessions</td>
<td>74</td>
<td>16%</td>
<td>33 (9.5)</td>
<td>66%</td>
</tr>
<tr>
<td>Rocha &amp; Queirós 2013</td>
<td>Clinically stable</td>
<td>MSCT</td>
<td>18 sessions</td>
<td>19</td>
<td>0%</td>
<td>39 (8.9)</td>
<td>84%</td>
<td>TAU</td>
<td>n.a.</td>
<td>16</td>
<td>0%</td>
<td>36 (8.7)</td>
<td>94%</td>
</tr>
<tr>
<td>Ross et al. 2013</td>
<td>Moderate to severe</td>
<td>MCT-JTC</td>
<td>45 minutes</td>
<td>17</td>
<td>0%</td>
<td>39 (10.2)</td>
<td>74%</td>
<td>TAU</td>
<td>n.a.</td>
<td>17</td>
<td>0%</td>
<td>36 (12.2)</td>
<td>71%</td>
</tr>
<tr>
<td>Balzan et al. 2014</td>
<td>Mild to moderate</td>
<td>MCT-T</td>
<td>1 hour</td>
<td>14</td>
<td>0%</td>
<td>38 (8.1)</td>
<td>78%</td>
<td>TAU</td>
<td>n.a.</td>
<td>14</td>
<td>0%</td>
<td>35 (8.7)</td>
<td>64%</td>
</tr>
<tr>
<td>Favrod et al. 2014</td>
<td>Mild to moderate</td>
<td>MCT</td>
<td>8 x MCT</td>
<td>26</td>
<td>8%</td>
<td>37 (9.8)</td>
<td>65%</td>
<td>TAU</td>
<td>n.a.</td>
<td>26</td>
<td>12%</td>
<td>37 (10.4)</td>
<td>65%</td>
</tr>
<tr>
<td>van Oosterhout et al. 2014</td>
<td>Moderate to severe</td>
<td>MCT</td>
<td>8 x MCT</td>
<td>75</td>
<td>31%</td>
<td>38 (11.1)</td>
<td>68%</td>
<td>TAU</td>
<td>n.a.</td>
<td>71</td>
<td>24%</td>
<td>37 (8.7)</td>
<td>71%</td>
</tr>
<tr>
<td>Kuokkanen et al. 2014</td>
<td>Minimal to mild</td>
<td>MCT</td>
<td>8 x MCT</td>
<td>10</td>
<td>20%</td>
<td>n.a.</td>
<td>n.a.</td>
<td>TAU</td>
<td>n.a.</td>
<td>10</td>
<td>0%</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Briki et al. 2014</td>
<td>Mild to severe</td>
<td>MCT</td>
<td>16 x MCT</td>
<td>35</td>
<td>17%</td>
<td>41 (8.1)</td>
<td>64%</td>
<td>SC</td>
<td>16 x ST</td>
<td>33</td>
<td>24%</td>
<td>41 (12.4)</td>
<td>68%</td>
</tr>
</tbody>
</table>

MCT = Meta-Cognitive Training; CBT = Cognitive Behavioural Therapy; TAU = Treatment as usual; MSCT = MCT plus Social Cognition training; MCT-T = 60 minutes single session focusing on data-gathering bias and confirmation bias modules; MCT-JTC = 45 minute single session partially based on MCT; SC = Supportive Counseling. * Moritz et al., 2011; Selected outcome was algorithm van der Gaag; ** Removed from analysis due to attending less than 8 sessions; *** Patients with scores 6 or 7 on PANSS paranoia/suspiciousness were excluded.
Heterogeneity is always a matter of concern in meta-analysis. Therefore, we evaluated whether the variability in the outcomes across the studies could be attributed to random sample error alone, or might be attributed to systematic factors, such as type of intervention. We tested heterogeneity with a Chi-square test and degrees of freedom (df) set at the number of primary studies in the meta-analysis minus one. We also report the $I^2$ statistic, which is easier to interpret: when $I^2 = 0\%$, 25%, 50% or 75%, then no, low, moderate or high heterogeneity, respectively, is assumed (Higgins et al., 2003).

Meta-analysis may be subject to publication bias. When publication bias was likely, then Duval and Tweedie’s trim and fill procedure was used; this yields an adjusted estimate of the pooled effect size after the publication bias has been taken into account (Duval & Tweedie, 2000).

**Sensitivity analyses**

The inclusion of studies was relatively liberal. High and low-quality studies with various types of statistical analyses and procedures to correct for unblinding, were selected. To examine the effects of study quality, we conducted additional sensitivity analyses in which we successively included high-quality, low-quality and blinded studies, as well as studies using proper intention-to-treat analysis. In order to correct for allegiance bias we used the criteria of the Researcher Allegiance Assessment Tool (Cuijpers et al., 2012) which accounts for the following criteria: 1) only one of the interventions was mentioned in the title; 2) one of the two interventions was explicitly mentioned as the main experimental intervention in the introduction section of the study; 3) one intervention was explicitly described as a control condition included to control for the non-specific components of the other therapy and 4) there was an explicit hypothesis that one comparison therapy was expected to be more effective than the other. When these criteria were applied none of the studies were non-alleged. Therefore we could not run analyses on allegiance bias.

**Results**

**Characteristics of the included studies**

Table 6.2 presents the results of the primary studies measuring positive symptoms analysis (upper part), delusions analysis (middle part) and data gathering bias (lower part).
## Table 6.2 Random effect-sizes, heterogeneity and publication bias in the main and sensitivity analyses

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Number of contrasts</th>
<th>Hedges’s g (95% CI)</th>
<th>Z</th>
<th>p-value of Z</th>
<th>Q (df)</th>
<th>p-value of Q</th>
<th>( I^2 )</th>
<th>Funnel plot</th>
<th>Trim and Fill Corrected Hedges’s g</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effects on positive symptoms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main with all studies</td>
<td>9</td>
<td>0.256 (-0.01–0.52)</td>
<td>1.927</td>
<td>0.054</td>
<td>12.507 (8)</td>
<td>0.130</td>
<td>36.0/ MOD</td>
<td>1 missing</td>
<td>0.207</td>
</tr>
<tr>
<td>High quality studies CTAM &gt; 65</td>
<td>2</td>
<td>0.279 (-0.18–0.74)</td>
<td>1.196</td>
<td>0.232</td>
<td>2.101 (1)</td>
<td>0.147</td>
<td>52.4/ MOD</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Low quality studies CTAM &lt; 65</td>
<td>7</td>
<td>0.224 (-0.12–0.60)</td>
<td>1.334</td>
<td>0.182</td>
<td>10.367 (6)</td>
<td>0.110</td>
<td>42.1/ MOD</td>
<td>0 missing</td>
<td>No correction</td>
</tr>
<tr>
<td>Blinded</td>
<td>4</td>
<td>0.359 (0.09–0.63)</td>
<td>2.570</td>
<td>0.010</td>
<td>3.921 (3)</td>
<td>0.270</td>
<td>23.5/ LOW</td>
<td>2 missing</td>
<td>0.223</td>
</tr>
<tr>
<td>Intention-to-treat analysis</td>
<td>1</td>
<td>0.098 (-0.22–0.42)</td>
<td>0.604</td>
<td>0.546</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td><strong>Effects on delusions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main with all studies</td>
<td>7</td>
<td>0.223 (-0.05–0.49)</td>
<td>1.630</td>
<td>0.103</td>
<td>11.306 (6)</td>
<td>0.079</td>
<td>46.9/ MOD</td>
<td>3 missing</td>
<td>0.034</td>
</tr>
<tr>
<td>High quality studies CTAM &gt; 65</td>
<td>3</td>
<td>0.108 (-0.30–0.52)</td>
<td>0.518</td>
<td>0.604</td>
<td>6.863 (2)</td>
<td>0.032</td>
<td>70.9/ HIGH</td>
<td>2 missing</td>
<td>-0.253</td>
</tr>
<tr>
<td>Low quality studies CTAM &lt; 65</td>
<td>4</td>
<td>0.387 (0.05–0.72)</td>
<td>2.257</td>
<td>0.024</td>
<td>1.643 (3)</td>
<td>0.650</td>
<td>0/ NO</td>
<td>1 missing</td>
<td>0.326</td>
</tr>
<tr>
<td>Blinded</td>
<td>5</td>
<td>0.174 (-0.12–0.47)</td>
<td>1.151</td>
<td>0.250</td>
<td>9.089 (4)</td>
<td>0.059</td>
<td>56.0/ HIGH</td>
<td>2 missing</td>
<td>0.028</td>
</tr>
<tr>
<td>Intention-to-treat analysis</td>
<td>2</td>
<td>-0.017 (-0.48–0.45)</td>
<td>-0.071</td>
<td>0.944</td>
<td>4.276 (1)</td>
<td>0.039</td>
<td>76.6/ HIGH</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td><strong>Effects on data-gathering bias</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main analysis (15–85 and 20–80)</td>
<td>3</td>
<td>0.307 (-0.16–0.77)</td>
<td>1.289</td>
<td>0.198</td>
<td>4.586 (2)</td>
<td>0.101</td>
<td>56.4/ MOD</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

\( Q = \) value for heterogeneity tested by Chi-square; \( I^2 = \) degree of heterogeneity; NO = No heterogeneity; MOD = Moderate heterogeneity; HIGH = High heterogeneity.
**Overall analysis on primary outcome measures**

**Positive symptoms**

**Overall analysis of the effects of MCT on positive symptoms**

The results on the positive symptoms are presented in Table 6.2 (upper panel) and Figure 6.2. The effect size (g = 0.26) showed a statistical tendency. But correction for publication bias reduced the effect size to non-significant (g = 0.21). The heterogeneity was moderate. Both the high and low-quality studies were non-significant. Four blinded studies had significant results (g = 0.36), but correction for publication bias reduced the effect size again to non-significant (g = 0.22). If proper intention-to-treat statistics were used (in one study only) the effect size was very small and non-significant (g = 0.10). For funnel plots, see Supplementary Figures S6.1 and S6.2.

### Effects on positive symptoms

<table>
<thead>
<tr>
<th>Study name</th>
<th>Hedges’s g</th>
<th>Standard error</th>
<th>Variance</th>
<th>Lower limit</th>
<th>Upper limit</th>
<th>Z-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kumar</td>
<td>0.919</td>
<td>0.500</td>
<td>0.250</td>
<td>-0.061</td>
<td>1.899</td>
<td>1.839</td>
<td>0.066</td>
</tr>
<tr>
<td>Moritz a</td>
<td>0.580</td>
<td>0.290</td>
<td>0.084</td>
<td>0.012</td>
<td>1.149</td>
<td>2.000</td>
<td>0.045</td>
</tr>
<tr>
<td>Moritz b</td>
<td>0.098</td>
<td>0.163</td>
<td>0.026</td>
<td>-0.220</td>
<td>0.417</td>
<td>0.604</td>
<td>0.546</td>
</tr>
<tr>
<td>Rocha &amp; Querios</td>
<td>-0.546</td>
<td>0.338</td>
<td>0.114</td>
<td>-1.209</td>
<td>0.116</td>
<td>-1.617</td>
<td>0.106</td>
</tr>
<tr>
<td>Naughton</td>
<td>0.355</td>
<td>0.448</td>
<td>0.200</td>
<td>-0.522</td>
<td>1.233</td>
<td>0.794</td>
<td>0.427</td>
</tr>
<tr>
<td>Fawrod</td>
<td>0.486</td>
<td>0.291</td>
<td>0.085</td>
<td>-0.085</td>
<td>1.057</td>
<td>1.668</td>
<td>0.095</td>
</tr>
<tr>
<td>Balzan</td>
<td>-0.054</td>
<td>0.367</td>
<td>0.135</td>
<td>-0.774</td>
<td>0.665</td>
<td>-0.148</td>
<td>0.883</td>
</tr>
<tr>
<td>Kuokkanen</td>
<td>0.070</td>
<td>0.452</td>
<td>0.204</td>
<td>-0.816</td>
<td>0.955</td>
<td>0.154</td>
<td>0.877</td>
</tr>
<tr>
<td>Briki</td>
<td>0.595</td>
<td>0.285</td>
<td>0.081</td>
<td>0.037</td>
<td>1.153</td>
<td>2.089</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>0.256</td>
<td>0.133</td>
<td>0.018</td>
<td>-0.004</td>
<td>0.516</td>
<td>1.927</td>
<td>0.054</td>
</tr>
</tbody>
</table>

**Figure 6.2** Forestplot of the effect on positive symptoms of psychosis.

a = Moritz et al., 2011; b = Moritz et al., 2013.

**Delusions**

**Overall analysis of the effects of MCT on delusions**

The results on the delusions are presented in the middle panel of Table 6.2 and Figure 6.3 and showed a non-significant effect size (g = 0.22), which was further reduced after correction for publication bias (g = 0.03). The level of heterogeneity was moderate. High-quality studies (g = 0.11), blinded studies (g = 0.17) and studies with intention-to-treat statistics (g = -0.02) were all non-significant. Just the low quality studies showed a significant effect. One by one removing of a single trial resulted in significance for the effect of MCT on delusions. This was the case if the van Oosterhout trial was removed: a trial with zero findings and high CTAM-score compared to the other trials. For funnel plots, see Supplementary Figures S6.3, S6.4, S6.5 and S6.6.
CHAPTER 6

Effects on delusions

<table>
<thead>
<tr>
<th>Study name</th>
<th>Hedges’s g</th>
<th>Standard error</th>
<th>Variance</th>
<th>Lower limit</th>
<th>Upper limit</th>
<th>Z-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kumar</td>
<td>0.836</td>
<td>0.495</td>
<td>0.245</td>
<td>-0.135</td>
<td>1.807</td>
<td>1.688</td>
<td>0.091</td>
</tr>
<tr>
<td>Moritz a</td>
<td>0.480</td>
<td>0.288</td>
<td>0.083</td>
<td>-0.085</td>
<td>1.045</td>
<td>1.665</td>
<td>0.096</td>
</tr>
<tr>
<td>Moritz b</td>
<td>0.221</td>
<td>0.163</td>
<td>0.027</td>
<td>-0.099</td>
<td>0.540</td>
<td>1.355</td>
<td>0.175</td>
</tr>
<tr>
<td>Favrod</td>
<td>0.507</td>
<td>0.292</td>
<td>0.085</td>
<td>-0.064</td>
<td>1.079</td>
<td>1.740</td>
<td>0.082</td>
</tr>
<tr>
<td>Otterhout</td>
<td>-0.253</td>
<td>0.161</td>
<td>0.026</td>
<td>-0.569</td>
<td>0.063</td>
<td>-1.571</td>
<td>0.116</td>
</tr>
<tr>
<td>Kuokkanen</td>
<td>0.304</td>
<td>0.431</td>
<td>0.186</td>
<td>-0.540</td>
<td>1.149</td>
<td>0.706</td>
<td>0.480</td>
</tr>
<tr>
<td>Briki</td>
<td>0.168</td>
<td>0.279</td>
<td>0.078</td>
<td>-0.378</td>
<td>0.715</td>
<td>0.604</td>
<td>0.546</td>
</tr>
<tr>
<td></td>
<td>0.223</td>
<td>0.137</td>
<td>0.019</td>
<td>-0.045</td>
<td>0.491</td>
<td>1.630</td>
<td>0.103</td>
</tr>
</tbody>
</table>

Figure 6.3 Forestplot on the effect on delusions.
a = Moritz et al., 2011; b = Moritz et al., 2013.

Data gathering bias

Overall analysis of the effects of MCT on delusions
The results on the data gathering bias are presented in the lower panel of Table 6.2 and Figure 6.4; results showed non-significant effect (g = 0.31). The level of heterogeneity was moderate.

Effects on data gathering

<table>
<thead>
<tr>
<th>Study name</th>
<th>Hedges’s g</th>
<th>Standard error</th>
<th>Variance</th>
<th>Lower limit</th>
<th>Upper limit</th>
<th>Z-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ross 2011</td>
<td>0.201</td>
<td>0.330</td>
<td>0.113</td>
<td>-0.457</td>
<td>0.059</td>
<td>0.600</td>
<td>0.549</td>
</tr>
<tr>
<td>Moritz 2013</td>
<td>0.050</td>
<td>0.163</td>
<td>0.026</td>
<td>-0.298</td>
<td>0.369</td>
<td>0.309</td>
<td>0.756</td>
</tr>
<tr>
<td>Rocha &amp; Queiroz 2013</td>
<td>0.056</td>
<td>0.340</td>
<td>0.115</td>
<td>0.191</td>
<td>1.522</td>
<td>2.522</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>0.307</td>
<td>0.236</td>
<td>0.057</td>
<td>-0.190</td>
<td>1.289</td>
<td>1.289</td>
<td>0.196</td>
</tr>
</tbody>
</table>

Figure 6.4 Forest plot on data-gathering bias.

Conclusions

Main findings
Currently, the evidence of this meta-analysis, does not support the efficacy of MCT for any of the outcomes selected. All main analyses on positive symptoms, delusions and data gathering bias, yielded non-significant effect sizes. Corrections for publication biases using
the Trim and Fill procedure further reduced the effect sizes. In general, the effect sizes were further reduced in high-quality studies, blinded studies, and studies using proper intention-to-treat analysis. The exception was the results of the blinded studies measuring positive symptoms, with results resembling the effects of CBT. Nevertheless, significance disappeared after correction for publication bias.

In almost all analyses there was a moderate to high level of heterogeneity, which makes it difficult to interpret the findings and increases the risk of bias. This raises the question to what extent other (methodological and clinical) trial characteristics may contribute to the effects in the various trials and whether positive or negative effects were exerted on the true effect sizes. Differences in types of patients, levels of delusional symptomatology at baseline, treatment dosage and lack of randomization and blindness are probably causing the heterogeneity in the results of the studies.

In our trial we found no effect on data gathering nor on delusions and a non-specific effect on positive symptoms in high quality studies (which disappeared after correction for publication bias). Regarding the data-gathering bias, only the Rocha Trial (Rocha and Queiros, 2013) found a significant effect, but also showed a worsening of positive symptoms instead of a reduction. The largest trial, from Moritz et al. (2013), found no effect on data gathering and only small non-significant effects on positive symptoms. Freeman et al. (2014) reported that only 24% of delusional patients showed a data-gathering bias and that this was associated with deficits in working memory, lower IQ, lower levels of tolerance for uncertainty, and lower worry. The data-gathering-bias was not associated with psychopathology in that study. Thus, data gathering might better be addressed by retraining working memory rather than by education. Also, the association between data-gathering bias and delusions is not very robust, as some studies found no such associations (Young & Bentall, 1997; McKay et al., 2007). This raises the question as to what extent data gathering (JTC) and delusions are causally linked, or whether JTC is an epiphenomenon related to psychosis and if making patients aware of cognitive biases and its negative consequences (aim of MCT / back-door approach) is necessary to achieve symptom reduction.

One of the most successful studies was conducted by Moritz et al. (2011a). This was in fact MCT plus CBT (mentioned as ‘MCT+’). The positive effect might be due to the effective CBT part of the intervention, rather than to the MCT part. The individual study was marginally significant on positive symptoms and showed a tendency on delusions. However, as this was not compared with CBT alone, the addition of MCT to the effective CBT cannot be evaluated at this moment. Furthermore it was observed that the developers of MCT have found positive results, but so far independent testing by other research groups has not indicated significant change in positive symptoms (with the exception of Briki et al. (2014) in a small study) nor in
delusions (with the exception of Favrod et al. (2014) in a small study). Moreover, in our study we found the latter studies to have relatively low CTAM-scores, reflecting lower methodological quality. More recent findings on MCT (Moritz et al., 2014) reported consolidation of delusion scores and consolidation of no effect on JTC. In the completer analysis with 40% drop-out the PANSS positive symptoms deteriorates in the control, probably due to an increase in hallucinations. The ITT analysis reported a group effect over all time moments and no group x follow-up interaction. We think these results are hard to interprete. Independent research indicating positive change is necessary for any treatment to be added to evidence based guidelines for routine care.

**Strengths and limitations**

The present study has several strengths and limitations. A strength is that separate meta-analyses were conducted on different outcome measures such as data gathering bias, delusions and positive symptoms.

Another strength was the statistical power to detect small to medium effect-sizes. There is little chance of making a type 2 error and incorrectly reject the hypothesis that MCT is efficacious.

At the same time the power is a limitation. Nevertheless, a cumulative analysis on the positive symptom and delusion outcomes showed that the effects stabilized at small and non-significant effect-sizes after five trials.

**General conclusions**

This is the first meta-analysis on MCT. Currently, we can state that the studies do not support a positive effect for MCT on positive symptoms, data gathering and delusions. The methodology of most studies was poor and sensitivity analyses with blinded studies, high quality studies, and studies that used intention-to-treat analyses reduced the effect-sizes even further. Also correction for publication bias reduced the effect-sizes considerably. Dissemination of MCT in routine care cannot be recommended at this moment. More rigorous research would be helpful in order to create enough statistical power to detect small effect sizes and to reduce heterogeneity.
References


Supplementary material

Supplementary Figure S6.1  Funnel plot of four blinded studies on positive symptoms with the effect size (Hedge’s g) on the horizontal and the standard error on the vertical axis.

Supplementary Figure S6.2  Funnel plot of all nine studies on positive symptoms with the effect size (Hedge’s g) on the horizontal and the standard error on the vertical axis.
Supplementary Figure S6.3  Funnel plot of five blinded studies on delusions with the effect size (Hedge's g) on the horizontal and the standard error on the vertical axis.

Supplementary Figure S6.4  Funnel plot of three high quality studies on delusions with the effect size (Hedge’s g) on the horizontal and the standard error on the vertical axis.
Supplementary Figure S6.5  Funnel plot of four low quality studies on delusions with the effect size (Hedge’s g) on the horizontal and the standard error on the vertical axis.

Supplementary Figure S6.6  Funnel plot of all seven studies on delusions with the effect size (Hedge’s g) on the horizontal and the standard error on the vertical axis.
Appendix 6.1

Letter to editor of Psychological Medicine in response to an invited commentary by Moritz et al.

In an invited commentary, the developers of the metacognitive training have challenged the results of the meta-analysis on the efficacy of metacognitive training on positive symptoms, delusions and data-gathering bias (van Oosterhout et al., 2015). The recommendation of this meta-analysis was not to implement metacognitive training (MCT) in routine care at this moment, because of a lack of evidence.

In their introduction Moritz et al. refer to another meta-analysis by Jiang and colleagues (2015) quoting significant effects for positive symptoms (as well as for delusions without the study by van Oosterhout and colleagues (2014)). Van Oosterhout is a negative study, but the largest randomized controlled trial (RCT) and indeed was rated as the study with the lowest risk of bias by Jiang and colleagues; it’s not clear why one would discard the best study to date? The developers could usefully be referred to the conclusion in the abstract of Jiang et al.: “The limited number of RCT trials, the variability of the method and time of the outcome evaluation, and methodological problems in the trials make it impossible to come to a conclusion about the effectiveness of MCT for schizophrenia. More randomized trials that use standardized outcome measures, that use intention-to-treat (ITT) analyses, and that follow-up participants at regular intervals after the intervention are needed to determine whether or not MCT should become a recommended adjunctive treatment for schizophrenia.” In our opinion this conclusion is the same as ours.

In their rebuttal the developers apply four headings with different arguments; these are addressed below.

Studies omitted
Moritz and colleagues note some studies were omitted; we reran our analyses including the appropriate studies. We added the data of Aghotor et al. (2010), Moritz, Kerstan et al. (2011) and Gaweda et al. (2015). We found the data in the meta-analysis of Eichner (2015): a PhD student of Moritz, who was so kind to provide the data via Research Gate. We discarded two non-randomized studies: an extremely negative study by Rocha & Queries (2013) and an extremely positive study by Erawati (2014).

It is not accurate to say that we omitted the So et al. study (2015) from our paper because it had not been published – or accepted for publication at the point our meta-analysis was accepted for publication. However, we have added this study in our reanalysis.
The results of the reanalysis demonstrate that there are some significant effects for positive symptoms (g = 0.32) and delusions (g = 0.31), but not for data-gathering (g = 0.11) if you consider all the studies. However, the findings from the high quality and intention to treat data are similar to our original results with no significant effects on delusions, positive symptoms or data-gathering.

**Heterogeneity and moderator analysis**
We have chosen to conduct some additional moderator analyses with the high quality studies and not the low quality studies as Moritz suggests. Van Oosterhout (2014) is the largest study with the lowest risk of bias, but is a negative study. It does not make sense to remove robust studies from an analysis only to remove heterogeneity.

**MCT+ is not CBT**
The authors wrote in the paper on MCT+ that “Individualized metacognitive therapy (MCT+) followed group sessions according to the general guidelines for CBT (e.g. Fowler et al., 1995).” As Fowler is one of the founding fathers of CBT for psychosis, this led us to conclude that MCT+ is based on CBT. In the same paper is stated “MCT...is grounded on the principles of CBT (Fowler et al. 1995) and basic research on cognitive biases in schizophrenia (for reviews, see Garety & Freeman, 1999; Freeman et al., 2007), as well as deficits in social cognition/theory of mind (Frith, 1994; Frith & Corcoran, 1996).” CBT in psychosis has included the work on cognitive biases for over a decade. The main difference seems to be that MCT is set up like a course targeting transfer of ‘cold cognitions’, while CBT by personalizing and experiencing the effects of biases in real life. In doing so, CBT is rather aimed at modifying ‘hot cognitions’.

**Evolution of MCT**
Moritz and colleagues suggested changing the inclusion criteria to patients without severe delusions. We have checked the baseline data of the studies on delusions. The three studies with the highest baseline delusion scores were So, Favrod and van Oosterhout. Taking these three studies, the effect-size is 0.49, while studies with the lowest delusion scores at baseline (3x Moritz, Briki (2014) and Gaweda) have a pooled effect-size of 0.25 on delusions. So the empirical evidence rather suggests that patients with highest baseline scores of paranoia in general benefit most from therapy.

In summary, we appreciate the opportunity to air these differences in opinion; but have to conclude that the best available evidence suggests that MCT is not yet at a stage to advocate its routine use. That is to say, we acknowledge Moritz and colleagues position that MCT is work in progress and this progress needs to be data driven. However, including less rigorous
evidence into a meta-analysis may offer a different opinion, but this is probably not the most robust scientific way forward.

References


