

# Planning performance in schizophrenia patients: a meta-analysis of the influence of task difficulty and clinical and sociodemographic variables

F. Knapp<sup>1\*</sup>, W. Viechtbauer<sup>2</sup>, R. Leonhart<sup>3</sup>, K. Nitschke<sup>3,4,5,6,7</sup> and C. P. Kaller<sup>4,5,6,7\*</sup>

<sup>1</sup>Kliniken Schmieder, Allensbach, Germany

<sup>2</sup>Department of Psychiatry and Neuropsychology, Maastricht University, Maastricht, The Netherlands

<sup>3</sup>Department of Psychology, University of Freiburg, Freiburg, Germany

<sup>4</sup>Department of Neurology, Medical Center, University of Freiburg, Freiburg, Germany

<sup>5</sup>Medical Faculty, University of Freiburg, Freiburg, Germany

<sup>6</sup>Freiburg Brain Imaging Center, University of Freiburg, Freiburg, Germany

<sup>7</sup>BrainLinks-BrainTools Cluster of Excellence, University of Freiburg, Freiburg, Germany

**Background.** Despite a large body of research on planning performance in adult schizophrenia patients, results of individual studies are equivocal, suggesting either no, moderate or severe planning deficits. This meta-analysis therefore aimed to quantify planning deficits in schizophrenia and to examine potential sources of the heterogeneity seen in the literature.

**Method.** The meta-analysis comprised outcomes of planning accuracy of 1377 schizophrenia patients and 1477 healthy controls from 31 different studies which assessed planning performance using tower tasks such as the Tower of London, the Tower of Hanoi and the Stockings of Cambridge. A meta-regression analysis was applied to assess the influence of potential moderator variables (i.e. sociodemographic and clinical variables as well as task difficulty).

**Results.** The findings indeed demonstrated a planning deficit in schizophrenia patients (mean effect size:  $\hat{\mu} = 0.67$ ; 95% confidence interval 0.56–0.78) that was moderated by task difficulty in terms of the minimum number of moves required for a solution. The results did not reveal any significant relationship between the extent of planning deficits and socio-demographic or clinical variables.

**Conclusions.** The current results provide first meta-analytic evidence for the commonly assumed impairments of planning performance in schizophrenia. Deficits are more likely to become manifest in problem items with higher demands on planning ahead, which may at least partly explain the heterogeneity of previous findings. As only a small fraction of studies reported coherent information on sample characteristics, future meta-analyses would benefit from more systematic reports on those variables.

Received 8 August 2016; Revised 27 January 2017; Accepted 31 January 2017; First published online 7 April 2017

**Key words:** Meta-analyses, schizophrenia, Stockings of Cambridge, Tower of Hanoi, Tower of London.

## Introduction

Cognitive impairment is a widely recognized finding in schizophrenia and has been established for chronic (Heinrichs & Zakzanis, 1998), first-episode (Mesholam-Gately *et al.* 2009) and drug-naïve patients (Fatouros-Bergman *et al.* 2014). Compared with healthy controls, the largest differences are reported for verbal episodic memory and processing speed (Dickinson *et al.* 2007; Mesholam-Gately *et al.* 2009; Palmer *et al.* 2009; Schaefer *et al.* 2013), but deficits also extend to

other domains such as attention, perception, language and visuo-spatial abilities (e.g. Heinrichs & Zakzanis, 1998; Fioravanti *et al.* 2005, 2012; Palmer *et al.* 2009).

Beyond these cognitive domains, deficits in executive functions and their association with clinical symptoms of schizophrenia have generated particular interest (e.g. Donohoe & Robertson, 2003). Executive functions constitute an umbrella term subsuming top-down cognitive processes related to the conscious control of behaviour (e.g. Lezak *et al.* 2004; Alvarez & Emory, 2006; Jurado & Rosselli, 2007). Although a uniform definition of executive functions still does not exist, there is broad consensus that working memory, inhibition and set shifting represent basic executive processes which contribute to high-order executive functions such as reasoning and problem solving (Miyake *et al.* 2000; Diamond, 2013). The ability to

---

\* Address for correspondence: F. Knapp and C. P. Kaller, Department of Neurology, University Medical Center Freiburg, Breisacher Str. 64, DE-79106 Freiburg, Germany.  
(Email: f.knapp@kliniken-schmieder.de) [F. K.]  
(Email: christoph.kaller@uniklinik-freiburg.de) [C. P. K.]

plan ahead future behaviour constitutes a prototypical example of high-level executive functioning, as it reflects the conscious selection of actions based on the anticipation of potential outcomes in relation to goals contingent upon current situational demands (Norman & Shallice, 1986). More specifically, planning is required for successful behaviour in many situations beyond everyday routine where known action schemata are either not applicable or suitable (Ward & Morris, 2005). In these instances, an appropriate and purposive behavioural sequence must be identified by mentally generating alternative sequences of interdependent actions and by evaluating the consequences of these anticipated actions in relation to goal attainment (Goel, 2002; Ward & Morris, 2005). Planning crucially depends on the integrity of the prefrontal cortex (Unterrainer & Owen, 2006; Nitschke *et al.* 2017), parts of which are known to be structurally and functionally affected in schizophrenia (Minzenberg *et al.* 2009).

Although planning deficits in schizophrenia patients have been reported previously (see Greenwood *et al.* 2011), a brief overview of the extant literature provides a highly inconsistent picture (see Sullivan *et al.* 2009): evidence on severe impairments in planning performance (e.g. Morice & Delahunty, 1996; Marczewski *et al.* 2001; Tyson *et al.* 2004) is in stark contrast to a considerable amount of studies reporting either no significant deficits (Krabbendam *et al.* 1999; Dichter *et al.* 2006; Feldmann *et al.* 2006; Greenwood *et al.* 2011; Asevedo *et al.* 2013) or only small decreases in planning performance of schizophrenia patients compared with matched controls (Badcock *et al.* 2005; Zhu *et al.* 2010). This heterogeneity is all the more surprising given that all of these studies assessed planning performance by using well-defined tower tasks such as the Tower of London (TOL), the Tower of Hanoi or one of their variants (Berg & Byrd, 2002).

In the light of these inconsistent findings on the extent of planning impairments in schizophrenia, the objectives of the present study were twofold. First, by conducting a comprehensive literature search and meta-analysis, we addressed the general question as to whether planning performance is indeed impaired in schizophrenia. More than 80 studies have examined planning in schizophrenia patients using tower tasks (see below), but – to the best of our knowledge – no quantitative evaluation of this literature in terms of a meta-analysis exists so far (but see Sullivan *et al.* 2009 for a first overview). Second, using moderator analyses, we aimed to investigate whether various task- and subject-related variables may account for the inconsistency of previous findings.

As planning tasks measure the ability to flexibly and consciously adapt behaviour to changing situational demands in a complex world, assessments of planning

performance may potentially constitute a valuable surrogate marker for a patient's ability to live an independent and autonomous life (see Holt *et al.* 2013). Understanding possible causes for the inconsistencies in the literature could thus be of utmost importance for future improvements in neuropsychological assessment in research and clinical contexts.

## Method

### Literature search

A comprehensive literature search using Medline, PsycINFO and ISI Web of Knowledge provided the basis of the present meta-analysis. Following the PRISMA guidelines ('Preferred Reporting Items for Systematic Reviews and Meta-Analyses'; Moher *et al.* 2009), the search was guided by a pre-specified review protocol stating the eligibility criteria. More specifically, manuscripts on tower tasks were selected using the search terms 'Tower of London', 'Tower of Hanoi', 'Tower of Toronto', 'CANTAB' and 'Stockings of Cambridge'. No specifications regarding the start date of publication were made. Out of the resulting 1224 references, 99 papers were identified that were published in English until January 2014 and that included samples of schizophrenia patients. After exclusion of reviews, case studies, unpublished dissertations and conference abstracts, a total of 84 studies remained. These studies were screened according to the following criteria: inclusion of independent patient and healthy control sample(s) at adult age and standardized application of a three-ball/disk tower version (TOL, Stockings of Cambridge and Tower of Hanoi) as a measure of planning performance.

Planning accuracy is commonly measured by calculating the number of problems that were perfectly solved in the minimum possible number of moves (see Berg & Byrd, 2002). As this turned out to be the most frequently reported outcome variable in the remaining studies (31 out of 44), the number of perfect solutions was used as the dependent variable in the present meta-analysis. A detailed flowchart of the selection process is provided in Fig. 1.

### Data extraction

From the selected studies, results and sample characteristics were extracted and entered in an Excel spreadsheet by F.K. and independently cross-checked by K.N. In the case of missing information on means and standard deviations, reported *t* or *F* statistics were converted to Hedges' *g* where applicable (Hedges & Olkin, 1985). In a few studies (Morris *et al.* 1995; Langdon *et al.* 2001; Joyce *et al.* 2002), the means and standard deviations

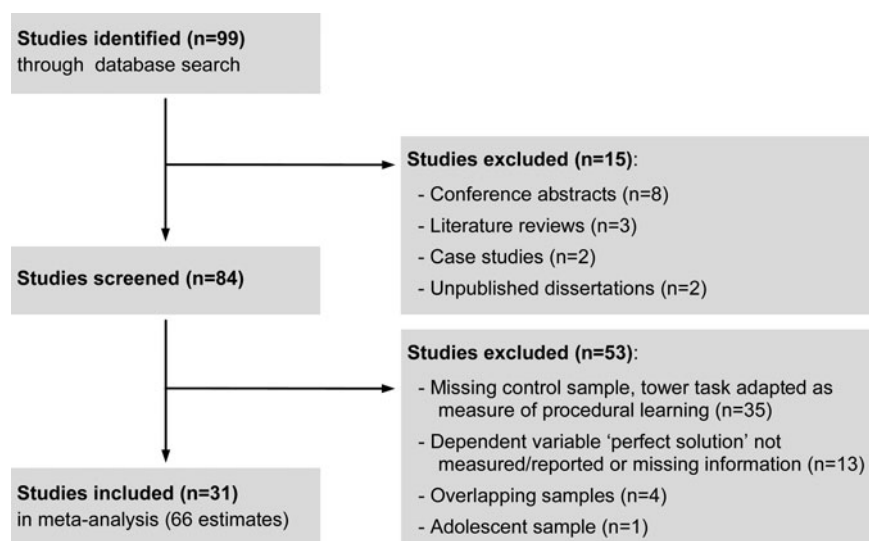


Fig. 1. Overview on the selection process.

needed to compute the  $g$  values were extracted from figures using graph digitizer software (DigitizeIT version 2.1; <http://www.digitizeit.de>; independent cross-check using WebPlotDigitizer version 3.9; Rohatgi, 2015). In 12 cases, the corresponding authors of the primary studies were contacted for further information. Of these, five authors provided the requested details, three provided additional data, and four did not respond.

### Statistical analyses

Following Hedges & Olkin (1985), reported means and standard deviations from each dataset were used to calculate standardized mean differences in terms of Hedges'  $g$ , thereby correcting for small-sample bias. The standardized mean difference reflects the difference in planning performance (or size of effect) between a given patient group and the respective group of healthy controls. Positive values indicate impaired planning performance in the patient group.

For several studies, multiple standardized mean differences were extracted (e.g. for different patient groups). The interdependency between these outcomes (e.g. different patient groups compared with one control group) was accounted for in the analyses by computing covariances between the dependent outcome measures and incorporating this information in the model (see online Supplementary material, section S1). A multilevel meta-analytic model was used including random effects at the study level and the effect-size level (Konstantopoulos, 2011; see also online Supplementary material, section S2, for more detailed information).

To evaluate potential sources of inconsistency between studies, the influence of potential moderator variables on the size of the (average) standardized mean difference was examined via meta-regression models by adding these potential moderators as fixed effects to the multilevel model (van Houwelingen *et al.* 2002). Based on available information, we hereby examined the role of task difficulty, sociodemographic variables [i.e. age, sex, intelligence quotient (IQ) and education level] and clinical variables (i.e. age at disease onset, disease duration, symptom severity and medication use; see Table 1).

For task difficulty, two analyses were conducted. The first analysis included all studies and outcomes, using one estimate of difficulty per outcome (in terms of the average minimum number of moves across problem items) as the predictor. In the second analysis, we restricted the model to data from studies reporting varying task difficulties (i.e. performance scores for different levels of minimum moves). The latter analysis thus focused on multiple estimates from the same sample of subjects, which partly circumvents the problem of comparing outcomes of different studies that varied in task difficulty.

The relevance of the demographic variables (i.e. age, sex, IQ and education level) was examined in two different ways in the analyses. First, we examined the impact of potential differences between patients and controls in the demographic variables on effect sizes. For this analysis, we used predictor variables that indicated the difference between the patient and the control group within each study. In particular, the standardized mean differences for age, IQ and education level (i.e. years of education) were computed

**Table 1.** Overview on the included studies and datasets

Study	Subsample	Demographic variables					Clinical variables					
		Sample size, <i>n</i> (SZ/HC)	Female, <i>n</i> (SZ/HC)	Age, years (SZ/HC)	Pre-morbid IQ (SZ/HC)	Level of education <sup>a</sup> (SZ/HC)	Disease		PANSS			
							Age at onset, years	Duration, years	Negative	Positive	General	Medication <sup>b</sup>
Andersen <i>et al.</i> (2013)		48/48	13/13	25.44/26.64		12.11/14.81		3.50	21.74	19.54	40.46	0.00
Asevedo <i>et al.</i> (2013) <sup>c</sup>		30/27	5/14	33.67/34.26		10.56/10.03			17.03	13.97	30.13	
Ayesa-Arriola <i>et al.</i> (2014) <sup>c</sup>	Male	86/97	0/0	29.03/28.40	36.69/40.78 <sup>d</sup>	9.63/10.40	27.92	1.11				0.00
	Female	75/62	75/62	35.97/30.00	38.96/39.76 <sup>d</sup>	11.27/11.32	34.80	1.17				0.00
Badcock <i>et al.</i> (2005)		24/33	5/7	32.83/34.67	101.42/108.27	11.50/13.10						
Braw <i>et al.</i> (2008) <sup>c</sup>	First-episode SZ	44/44	10/17	24.01/25.55		11.93/13.78	22.00	2.02	21.97	13.64	44.04	244.76
	Multi-episode SZ	38/44	15/17	25.82/25.55		12.26/13.78	20.19	5.34	23.06	15.27	44.18	281.90
Braw <i>et al.</i> (2012) <sup>c</sup>	Symptomatic	39/37	10/14	27.33/28.60		11.97/12.69	21.49		25.85	17.26	49.67	330.03
	Positive remission	27/37	6/14	29.82/28.60		12.44/12.69	21.93		21.96	10.96	37.56	308.58
	Negative remission	15/37	5/14	26.80/28.60		12.07/12.69	21.67		16.40	17.87	42.20	254.02
	Full remission	20/37	7/14	29.10/28.60		11.85/12.69	22.38		14.85	11.65	33.30	239.45
Bustini <i>et al.</i> (1999)		28/28	8/10	31.50/29.20		12.10/13.20	23.60	7.90				664.20
Dichter <i>et al.</i> (2006)		13/10	4/N.A.	28.20/29.70		13.00/15.30	21.70	5.90	14.40	12.00	27.00	462.00
Elliott <i>et al.</i> (1998)		12/12	1/1	43.40/40.00	110.40/111.40			20.00				870.00
Fagerlund <i>et al.</i> (2004)		25/25							19.45	20.00	30.00	0.00
Feldmann <i>et al.</i> (2006) <sup>c</sup>		22/20	7/11	34.90/31.00	106.60/116.70	14.00/16.60	24.80	10.40				524.80
Goldberg <i>et al.</i> (1990)		13/15		31.00/30.00		13.00/15.00	22.00	9.00				
Greenwood <i>et al.</i> (2011)	Psychomotor poverty symptoms	27/28	N.A./3	35.20/33.10	97.30/110.00	12.80/14.50		7.83				
	Disorganized symptoms	28/28	3.50/3	36.20/33.10	96.80/110.00	12.10/14.50		11.90				
Hilti <i>et al.</i> (2010)		26/33	N.A./9	22.00/23.20	93.00/101.00	9.50/12.70			14.50	17.10	35.50	0.00
Huddy <i>et al.</i> (2007)		20/20	8/10	23.50/29.00	90.10/94.40							
Joyce <i>et al.</i> (2002)		135/81	28.50/32	25.74/26.12	99.67/104.64							
Kontis <i>et al.</i> (2013) <sup>c</sup>		77/55	27.50/25	42.26/43.69		11.01/13.46	26.17	16.55	20.26	19.36	36.39	928.72
Krabbendam <i>et al.</i> (1999)		24/17	4/2	30.70/30.60		3.90/4.20 <sup>e</sup>	24.90	8.20				
Langdon <i>et al.</i> (2001)		32/24	14/12	37.31/34.54		11.91/12.96	24.30	12.90				
Langdon <i>et al.</i> (2002) <sup>c</sup>		25/20					23.20	9.60				
Marczewski <i>et al.</i> (2001)		15/15	6/6	29.93/30.47		10.60/10.67		8.07				296.30
Morice & Delahunty (1996)		17/17	7/12	32.41/34.00	107.80/112.70		23.10	9.40				294.00

Morris et al. (1995)	30/27	5/N.A.	32.10/30.30	98.79/100.65	12.70/13.50	24.00	8.20	646.30
Okruszek & Rutkowska (2013)	30/30	0/0	45.67/42.97					
Pantelis et al. (1997)	36/31	7/13	48.31/47.48	97.16/101.27		20.69	27.61	1401.10
Saleem et al. (2013) <sup>c</sup>	20/15	4/3	26.50/23.80	94.70/98.10	11.90/12.40	17.00	12.10	28.20
Shum et al. (2004)	60/60	18/18	31.10/30.60		10.20/10.50			
Tenjin et al. (2012)	20/340	10/N.A.	26.10/N.A.		13.80/N.A.	29.50	23.80	56.90
	10/340	5/N.A.	26.90/N.A.		13.20/N.A.	18.20	15.40	37.30
Tsuchimine et al. (2013)	56/68	29/16	42.30/36.40	97.90/101.70	12.60/14.00	28.00	14.90	536.60
	47/62	22/16	39.10/36.40	96.40/100.30	13.50/14.00	24.50	10.90	532.40
	15/19	7/4	30.10/38.90	95.20/104.50	13.50/14.20	24.20	6.50	872.60
Tyson et al. (2004)	28/17	10/N.A.	33.85/39.40	101.17/106.17				
Zhu et al. (2010) <sup>c</sup>	40/40	20/22	22.80/24.40		13.29/14.08	20.97	1.29	33.83

PANSS, Positive and Negative Syndrome Scale; SZ, schizophrenia; HC, healthy control; IQ, intelligence quotient; N.A., not available.

<sup>a</sup> Years of education.

<sup>b</sup> Chlorpromazine equivalent dose.

<sup>c</sup> Clarification and additional data obtained through personal communication.

<sup>d</sup> Wechsler Adult Intelligence Scale, third edition (WAIS-III) vocabulary raw scores.

<sup>e</sup> Measured on an eight-point scale.

between each pair of patient and control group by subtracting the value of the patient group from the value of the control group. For sex, we computed the difference in the proportion of females between each pair. Second, we examined the overall impact of demographic variables on effect sizes by including the mean age, sex (proportion of females), IQ and education level for each study (i.e. for patient and control groups combined) as predictors.

We further examined potential moderator effects of various clinical variables (i.e. age at disease onset, disease duration, symptom severity and medication use). For symptom severity, only those studies were included that reported negative, positive, general and total symptom scores according to the Positive and Negative Symptom Scale [i.e. Positive and Negative Syndrome Scale (PANSS); Kay et al. 1989].

Funnel plot asymmetry as a possible indicator for publication bias was examined visually (by plotting the observed standardized mean differences against their standard errors) and by adding the sampling variances of the standardized mean differences as a predictor to the multilevel model (which essentially models a quadratic relationship between the standardized mean differences and the standard errors). The intercept of this model was also used as an estimate of the 'genuine' effect devoid of publication selection bias [analogous to the 'precision-effect estimate with SE' (PEESE) method; see Stanley & Doucouliagos, 2014].

All models were fitted using restricted maximum likelihood estimation (van Houwelingen et al. 2002). Analyses were carried out in R (version 3.2.3; R Core Team, 2015) using the metafor package (version 1.9.8; Viechtbauer, 2010).

## Results

Literature search and subsequent selection (Fig. 1) resulted in 66 effect size estimates extracted from 31 studies. The majority of studies (i.e. 17) provided only one estimate of the standardized mean difference, whereas multiple estimates were extracted from 14 studies (online Supplementary material, section S1). In sum, the included studies comprised the data of 1377 schizophrenia patients and 1477 healthy controls. Diagnoses of schizophrenia were based on the Diagnostic and Statistical Manual of Mental Disorders, third edition (DSM-III), DSM-III, revised (DSM-III-R), DSM, fourth edition (DSM-IV), DSM-IV, text revision (DSM-IV-TR) or the International Classification of Diseases (ICD)-10 in all but one study (Elliott et al. 1998). In 29 studies, planning performance was assessed using variants of the TOL, whereas variants of the Tower of Hanoi were applied



in the remaining two studies. For more detailed information, please refer to the overview on all included studies and datasets provided in Table 1.

### Average standardized mean difference in planning performance

The observed standardized mean differences in planning performance ranged from  $-0.06$  to  $2.58$  (see histogram in Fig. 2a), with a mean difference of  $0.71$  (median,  $0.64$ ) (Fig. 3). The average standardized mean difference as estimated in the multilevel model was  $\hat{\mu} = 0.67$  with the 95% confidence interval (CI) ranging from  $0.56$  to  $0.78$ . The true effects appeared to be heterogeneous ( $Q = 147.63$ , degrees of freedom =  $65$ ,  $p < 0.0001$ ). The larger part of the heterogeneity in the true effects was attributable to differences in the true effects within studies ( $\hat{\sigma}_2^2 = 0.0509$ ), but there was also noteworthy between-study heterogeneity ( $\hat{\sigma}_1^2 = 0.0137$ ). Identifiability of the variance components was ensured by inspection of corresponding profile likelihood plots (see online Supplementary material, section S3).

The 95% CI of  $0.56$ – $0.78$  denotes the interval within which the average standardized mean difference can be expected, but it does not indicate where the true standardized mean difference in any particular study would be expected to be. To estimate this true standardized mean difference in a particular study, we thus computed the 95% prediction interval, which was found to range between  $0.16$  and  $1.18$ . While the interval encompasses effects that could be considered to range from very small to very large, it does not include the value  $0$ . This suggests that the group difference is expected to be present in any particular single study with high probability, albeit its exact magnitude being uncertain.

### Robustness of the results

An examination of the standardized residuals revealed three reported effects with relatively large absolute values of  $\geq \pm 2$  (Morice & Delahunty, 1996; Langdon *et al.* 2001; Marczewski *et al.* 2001). In fact, these studies yielded the smallest/largest standardized mean differences in the dataset (i.e.  $-0.06$ ,  $1.89$  and  $2.64$ , respectively). Removing these estimates led to a slight decrease in the estimated average effect ( $\hat{\mu} = 0.64$ , 95% CI  $0.54$ – $0.73$ ), but not to a change in the conclusions.

Two studies (Goldberg *et al.* 1990; Bustini *et al.* 1999) measured performance differences using the Tower of Hanoi task, which is similar but not identical to the TOL task<sup>†</sup> used in the remaining studies. After

removing these two studies, model estimation yielded virtually identical results ( $\hat{\mu} = 0.66$ , 95% CI  $0.56$ – $0.77$ ), thus indicating that the results were not influenced by the type of tower task used.

### Publication bias

According to the funnel plot (Fig. 2b), some small studies with estimates close to zero are possibly missing. As a formal test for funnel plot asymmetry, the sampling variance of the standardized mean differences was added as a predictor to the model, which yielded evidence of a significant relationship ( $p < 0.001$ ) and hence indicates possible publication bias. As discussed by Stanley & Doucouliagos (2014), the intercept of this model can be used to estimate the effect free of publication selection bias. As shown in Fig. 2b (solid line), this yielded an estimate of  $0.38$  (95% CI  $0.20$ – $0.56$ ), which can be regarded as substantially lower than the estimate of  $0.67$  obtained in the main analysis, but still significantly different from zero.

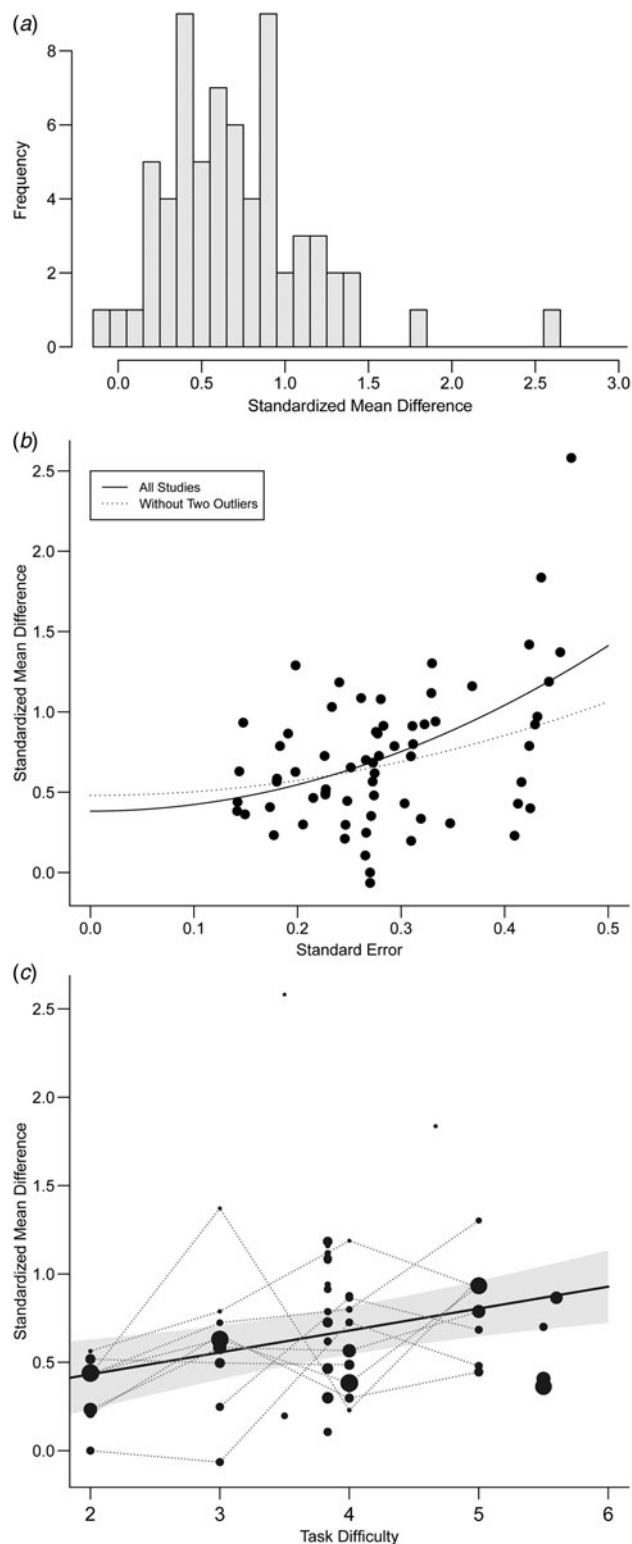
These findings appeared to be driven to a large extent by the two largest positive estimates (i.e.  $1.83$  from Marczewski *et al.* 2001 and  $2.58$  from Morice & Delahunty, 1996). Removing these two studies reduced the relationship between the standardized mean differences and the sampling variances to just below significance ( $p = 0.06$ ) and resulted in a higher intercept estimate of  $0.48$  (95% CI  $0.30$ – $0.66$ ; see Fig. 2b, dotted line).

### Moderator analyses

#### Effects of task difficulty

Information of task difficulty (in terms of the minimum number of moves required for the optimal solution) was available for 57 out of the 66 effect size estimates. Meta-regression analysis yielded a significant relationship between overall task difficulty and the size of the (average) effect ( $b = 0.124$ , s.e. =  $0.038$ ,  $p = 0.001$ ; online Supplementary material, Supplementary Table S4-1; see also Fig. 2c), with effect sizes increasing with higher task difficulty. That is, an increase in a problem set's difficulty of one minimum move amplified the estimated effect size of planning differences between groups of schizophrenia patients and healthy controls by  $0.12$  units. The analysis showed that there was a 95% chance of the true standardized mean difference in a particular study being larger than  $0$  if its task difficulty was three moves or higher (Table 2). Moreover, a medium effect size (i.e. a standardized mean difference above  $0.5$ ) was reliably achieved if the task difficulty comprised at least four minimum moves ( $\hat{\mu} = 0.68$ , 95% CI  $0.54$ – $0.82$ ).

<sup>†</sup> The notes appear after the main text.



**Fig. 2.** (a) Histogram of the standardized mean differences across all included datasets. (b) (Funnel) plot of the standard errors against the observed standardized mean differences. The regression line is based on the model including the sampling variances as a predictor (as a formal test for funnel plot asymmetry). (c) Scatterplot of task difficulty against the effect size estimates (drawn inversely proportional to their standard errors). The grey shaded area corresponds to a 95% confidence interval band for the predicted effect size as a function of task difficulty; the solid black line corresponds to the resulting regression line. Multiple estimates extracted from the same study based on the same tower task but with different difficulties are connected by dotted lines.

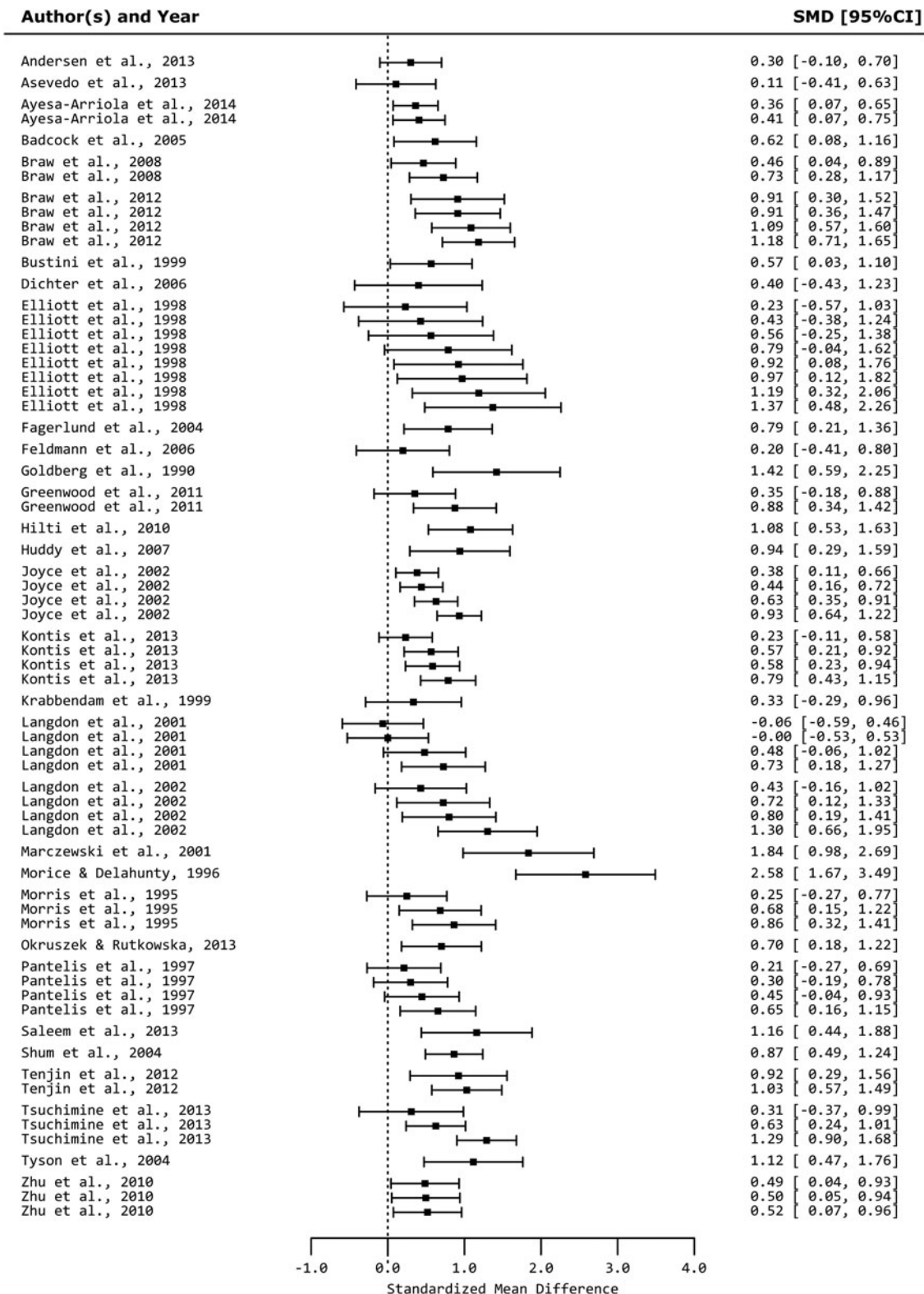


Fig. 3. Forest plot of the standardized mean difference (SMD) between schizophrenia patients and healthy controls across all included datasets. Note that the 95% confidence intervals (CIs) include the value of 0 in a substantial part of the datasets.



**Table 2.** Effect of task difficulty on effect size

Difficulty <sup>a</sup>	Pred	s.e.	ci.lb	ci.ub	pi.lb	pi.up
2	0.43	0.10	0.23	0.63	-0.14	1.00
2.5	0.49	0.09	0.32	0.67	-0.07	1.05
3	0.56	0.08	0.40	0.71	0.00	1.11
3.5	0.62	0.07	0.48	0.76	0.07	1.17
4	0.68	0.07	0.54	0.82	0.13	1.23
4.5	0.74	0.07	0.60	0.88	0.19	1.29
5	0.80	0.08	0.65	0.96	0.25	1.36
5.5	0.87	0.09	0.69	1.05	0.30	1.43

pred, Predicted average standardized mean difference; s.e., standard error; ci.lb, lower-bound confidence interval; ci.ub, upper-bound confidence interval; pi.lb, lower-bound prediction interval; pi.up, upper-bound prediction interval.

<sup>a</sup> Number of minimum moves for a perfect solution.

Differences in task difficulty across studies (or any other potential moderator) could be easily confounded with unknown third variables, leading to spurious relationships – a general and well-known problem with meta-regression analyses (Thompson & Higgins, 2002). This problem can be diminished by analysing studies that provide multiple estimates from the same sample of subjects. We therefore repeated the analysis including only those eight datasets reporting performance differences at various levels of task difficulty (i.e. the points in Fig. 2c connected by the dotted lines;  $k=34$  effect size estimates). The relationship between task difficulty and outcome was of comparable magnitude ( $b=0.137$ ,  $s.e.=0.039$ ) and remained significant ( $p<0.001$ ; online Supplementary material, Supplementary Table S4-1), providing additional support for the hypothesis that the difference in planning performance between groups of schizophrenia patients and controls increased as studies/datasets applied more difficult problem items.

#### Effects of sociodemographic and clinical variables

Ideally, the groups being compared in the individual studies should be as comparable as possible, except for patient status. To examine the respective fits, information about the age, sex, IQ and level of education was coded for each group where available. Based on this information, the standardized mean differences for age, IQ and education were computed for each pair of groups being compared. For sex, we computed the difference in the proportion of females within each pair of groups. Summary statistics are provided in online Supplementary material (Supplementary Table S4-2). While on average groups did not differ noticeably with respect to age and sex, the healthy

control groups had on average higher IQ scores and a higher education level (by approximately half of a standard deviation). However, when using the difference scores between patients and controls for age, sex, IQ and education as predictors in additional meta-regression analyses, none of these variables was found to be significantly related to the size of the group difference in planning performance (all  $p \geq 0.31$ ; online Supplementary material, Supplementary Table S4-3).

Similarly, when using mean age, proportion of females, mean IQ and mean educational level for each patient and control group pair combined as potential moderator variables, we did not find any significant relationships (all  $p \geq 0.53$ ; online Supplementary material, Supplementary Table S4-4).

Finally, patient groups differed with respect to various clinical variables, including mean age at disease onset, disease duration, symptom severity (PANSS negative, positive, general and total symptom scores), and medication use (measured in terms of the equivalent mean daily doses of chlorpromazine per 100 mg). In additional meta-regression models, we therefore examined to what extent these variables were related to the outcome. Again, none of these variables was found to be related to the size of the performance difference between groups (all  $p \geq 0.20$ ; online Supplementary material, Supplementary Table S4-5).

#### Discussion

In light of the hitherto heterogeneous findings from individual studies (see the forest plot in Fig. 3), the present meta-analysis resolves this inconsistency by demonstrating that schizophrenia is indeed associated with impaired planning performance. Assuming that planning performance is normally distributed within patients and controls (see Kaller *et al.* 2016), the standardized mean difference between groups of 0.67 implies that a randomly chosen healthy control subject would outperform a randomly chosen patient with 68.3% probability (95% CI 65.5–70.9%)<sup>2</sup>. The present results further demonstrate that these impairments in schizophrenia patients are amplified by increased planning demands in terms of the minimum number of moves to a solution, thus suggesting that the performance deficit of schizophrenia patients on tower tasks is indeed reflecting a specific planning deficit rather than unspecific cognitive impairment.

#### Effects of sociodemographic and clinical variables

The present results point to an overall planning deficit in schizophrenia, but further reveal considerable heterogeneity (Figs 2a and 3) which has also been

observed in meta-analyses of impairments of schizophrenia patients in other cognitive domains (Fioravanti *et al.* 2005, 2012; Forbes *et al.* 2009; Mesholam-Gately *et al.* 2009; Knowles *et al.* 2010). Regarding planning performance, various authors have attempted to explain this heterogeneity by linking planning deficits to specific clinical variables. For instance, Greenwood *et al.* (2011) have suggested that planning performance is specifically affected in patients with (positive) disorganization symptoms compared with patients with (negative) psychomotor poverty symptoms. Morice & Delahunty (1996) also found an association between TOL performance and positive symptoms. In contrast, Braw *et al.* (2012) showed that only patients in positive but not in negative symptomatic remission exhibited planning deficits, underscoring the general role of (persisting) negative symptoms for cognitive dysfunction (Ventura *et al.* 2009). Thus, while specific patient characteristics seem to exert a differential impact on planning performance, the pattern behind these partly controversial findings is still not understood. In this respect, meta-regression analysis would constitute a powerful tool for unveiling moderating effects of clinical and/or sociodemographic characteristics that potentially drive these differences between studies (see Forbes *et al.* 2009; Knowles *et al.* 2010; Fioravanti *et al.* 2012).

However, the present attempt to elucidate the moderating effect of sociodemographic and clinical variables on planning impairments was significantly hampered by the lack of necessary information and the low consistency of which details were reported across the studies. More specifically, from the initially identified 99 studies, only 31 studies provided sufficient information for the meta-analysis. An even smaller number of eight to 28 studies could be included in the subsequent meta-regression analyses of potential sociodemographic (i.e. age, sex, IQ and education) and clinical moderator variables (i.e. age at disease onset, disease duration, severity of positive and negative symptoms, medication; see online Supplementary material, section S4). This problem of missing information clearly limits the generalizability of the present results. In particular, it cannot be ruled out that possible effects of sociodemographic and/or clinical characteristics on the existence or severity of planning deficits have remained undetected due to limited statistical power. Notably, the importance of increasing the degree of overlapping information on patient characteristics across studies has already been pointed out in previous meta-analyses on cognitive impairment in schizophrenia patients (Mesholam-Gately *et al.* 2009; Fioravanti *et al.* 2012).

### Effects of task difficulty

As a second key finding, the present meta-regression analyses revealed that planning deficits of schizophrenia patients were exacerbated by increasing task demands (see Morice & Delahunty, 1996). Notably, two-move TOL problems do not require planning ahead at all but can be solved by simple perceptual matching strategies (Owen, 2005), whereas three-move TOL problems require to plan ahead only one move at maximum. This skill is reliably mastered, for instance, by neurotypical children at the age of 6 years (Kaller *et al.* 2008; McCormack & Atance, 2011; Unterrainer *et al.* 2013). Deficits in these very easy types of problems may hence be indicative of more general impairments in lower-order cognitive processes such as selective attention, working memory or inhibition rather than of specific planning impairments. Planning impairments in more difficult problems may also be partly driven by increasing task demands on lower-order cognitive processes. However, the finding that differences between schizophrenia patients and healthy controls were larger for problems with higher planning demands (i.e. with four minimum moves and more) than for the easy two- and three-move problems attests to a genuine planning deficit in schizophrenia.

The present moderator analyses imply that planning demands of TOL problems are mainly determined by the minimum number of moves. However, recent research has highlighted the impact of additional structural problem parameters such as search depth and goal hierarchy on planning performance (e.g. Ward & Allport, 1997; Carder *et al.* 2004; Kaller *et al.* 2004; Unterrainer *et al.* 2005; Newman & Pittman, 2007; Berg *et al.* 2010; Nitschke *et al.* 2012; see Kaller *et al.* 2011a, for an overview). These additional determinants of planning difficulty differentially affect planning performance in childhood and older age (Kaller *et al.* 2008; Unterrainer *et al.* 2013, 2015a; Köstering *et al.* 2014), in patients with various neurodevelopmental or neurological pathologies (McKinlay *et al.* 2008; Köstering *et al.* 2012; Rainville *et al.* 2012; Unterrainer *et al.* 2015b) and further elicit differential patterns of brain activation (Newman *et al.* 2009; Kaller *et al.* 2011b; Ruh *et al.* 2012). Instead of a global deficit, planning impairments in schizophrenia might hence be differentially associated with specific planning demands, for instance, on the depth and/or breadth of searching ahead. As these additional structural problem parameters are seldom systematically accounted for in existing TOL versions, the extant heterogeneity of findings on planning impairments in schizophrenia across studies (Figs 2a and 3) may at least to some

extent be attributable to possible differences in the structural properties of the problem sets applied, thus taxing different specific planning demands across studies.

Not only the insufficient standardization of the TOL as a measurement instrument has been criticized (Kaller et al. 2011a), but also the reliability (Humes et al. 1997; Lowe & Rabbitt, 1998; Syväoja et al. 2015) and validity (Kafer & Hunter, 1997) of some TOL versions has been questioned, which may have likewise contributed to the heterogeneous findings (but see Culbertson & Zillmer, 1998a, b; Schnirman et al. 1998; Kaller et al. 2012, 2016; Köstering et al. 2015a, b; Debelak et al. 2016; Tunstall et al. 2016 for suggestions of TOL versions with adequate psychometric properties).

### Limitations

Several limitations need to be taken into account. The results indicate possible publication bias, with small studies yielding estimates close to zero possibly missing. While the estimated standardized mean difference of 0.67 may be an overestimation, the most conservative estimate of 0.38 still indicates a small to medium effect size.

As stated earlier, the lack of overlap in reported sample characteristics between studies severely constrained the options for further analyses and the potential for identifying variables driving the observed heterogeneity of effect sizes. Limited statistical power constitutes one issue, but meta-regression analysis in the presence of heterogeneity and a low number of observations may also result in inflated rates of false-positive findings (Higgins & Thompson, 2004). Furthermore, potential interdependencies of moderating effects in terms of higher-order interactions (e.g. minimum moves by symptom severity) could not be tested due to the small numbers of reports providing sufficient information. Finally, the present results are focused on accuracy scores and the number of problems optimally solved in the minimum number of moves as the most frequently reported outcome (Berg & Byrd, 2002; Sullivan et al. 2009) but cannot be extrapolated to other outcome variables (e.g. pre-planning and execution times).

### Conclusions

Taken together, the present results advocate using psychometrically sound TOL versions with a graded difficulty of at least four minimum moves for reliably identifying planning impairments in schizophrenia patients. Future meta-regression analyses, particularly on the impact of task-specific as well as clinical and

sociodemographic variables on planning impairments in schizophrenia, are needed to further investigate the heterogeneity of effects. Such analyses, however, require more comprehensive reports of sample and task characteristics than is currently provided by studies on planning performance in schizophrenia. Notwithstanding this heterogeneity in the extent of effects, it can be concluded that schizophrenia indeed incurs a significant and genuine deficit in the ability to plan ahead one's own actions.

### Notes

<sup>1</sup> Note that the Stockings of Cambridge constitute a variant of the TOL that uses the same rules and the same problem space, but has a different physical appearance (balls are placed in sockets instead of on rods).

<sup>2</sup> Uncertainty about this value is larger for a given particular study, as reflected by the 95% prediction interval (54.6–79.8%).

### Supplementary material

The supplementary material for this article can be found at <https://doi.org/10.1017/S0033291717000459>

### Acknowledgements

This research was partly supported by the BrainLinks-BrainTools Cluster of Excellence funded by the German Research Foundation (DFG; grant no. EXC 1086). The authors thank Dr Benjamin Rahm and Dr Lena Schumacher for their valuable comments on a previous version of the paper.

### Declaration of Interest

None.

### References

- Alvarez JA, Emory E (2006). Executive function and the frontal lobes: a meta-analytic review. *Neuropsychology Review* **16**, 17–42.
- Andersen R, Fagerlund B, Rasmussen H, Ebdrup BH, Aggermaes B, Gade A, Oranje B, Glenthøj B (2013). The influence of impaired processing speed on cognition in first-episode antipsychotic-naïve schizophrenic patients. *European Psychiatry: the Journal of the Association of European Psychiatrists* **28**, 332–339.
- Asevedo E, Gadelha A, Noto C, Mansur RB, Zugman A, Belangero SIN, Berberian AA, Scarpato BS, Leclerc E, Teixeira AL, Gama CS, Bressan RA, Brietzke E (2013). Impact of peripheral levels of chemokines, BDNF and oxidative markers on cognition in individuals with schizophrenia. *Journal of Psychiatric Research* **47**, 1376–1382.

- Ayesa-Arriola R, Rodriguez-Sanchez JM, Gomez-Ruiz E, Roiz-Santiañez R, Reeves LL, Crespo-Facorro B (2014). No sex differences in neuropsychological performance in first episode psychosis patients. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* **48**, 149–154.
- Badcock JC, Michiel PT, Rock D (2005). Spatial working memory and planning ability: contrasts between schizophrenia and bipolar I disorder. *Cortex* **41**, 753–763.
- Berg K, Byrd D (2002). The Tower of London spatial problem-solving task: enhancing clinical and research implementation. *Journal of Clinical and Experimental Neuropsychology* **24**, 586–604.
- Berg WK, Byrd DL, McNamara JPH, Case K (2010). Deconstructing the tower: parameters and predictors of problem difficulty on the Tower of London task. *Brain and Cognition* **72**, 472–482.
- Braw Y, Benozio A, Levkovitz Y (2012). Executive functioning during full and partial remission (positive and negative symptomatic remission) of schizophrenia. *Schizophrenia Research* **142**, 122–128.
- Braw Y, Bloch Y, Mendelovich S, Ratzoni G, Gal G, Harari H, Tripto A, Levkovitz Y (2008). Cognition in young schizophrenia outpatients: comparison of first-episode with multipisode patients. *Schizophrenia Bulletin* **34**, 544–554.
- Bustini M, Stratta P, Daneluzzo E, Pollice R, Prosperini P, Rossi A (1999). Tower of Hanoi and WCST performance in schizophrenia: problem-solving capacity and clinical correlates. *Journal of Psychiatric Research* **33**, 285–290.
- Carder H, Handley S, Perfect T (2004). Deconstructing the Tower of London: alternative moves and conflict resolution as predictors of task performance. *Quarterly Journal of Experimental Psychology* **57**, 1459–1483.
- Culbertson WC, Zillmer EA (1998a). The construct validity of the Tower of London DX as a measure of the executive functioning of ADHD children. *Assessment* **5**, 215–226.
- Culbertson WC, Zillmer EA (1998b). The Tower of London<sup>DX</sup>: a standardized approach to assessing executive functioning in children. *Archives of Clinical Neuropsychology* **13**, 285–301.
- Debelak R, Egle J, Köstering L, Kaller CP (2016). Assessment of planning ability: psychometric analyses on the unidimensionality and construct validity of the Tower of London Task (TOL-F). *Neuropsychology* **30**, 346–360.
- Diamond A (2013). Executive functions. *Annual Review of Psychology* **64**, 135–168.
- Dichter GS, van der Stelt O, Boch JL, Belger A (2006). Relations among intelligence, executive function, and P300 event related potentials in schizophrenia. *Journal of Nervous and Mental Disease* **194**, 179–187.
- Dickinson D, Ramsey M, Gold J (2007). Overlooking the obvious: a meta-analytic comparison of digit symbol coding tasks and other cognitive measures in schizophrenia. *Archives of General Psychiatry* **64**, 532–542.
- Donohoe G, Robertson IH (2003). Can specific deficits in executive functioning explain the negative symptoms of schizophrenia? A review. *Neurocase* **9**, 97–108.
- Elliott R, McKenna PJ, Robbins TW, Sahakian BI (1998). Specific neuropsychological deficits in schizophrenic patients with preserved intellectual function. *Cognitive Neuropsychiatry* **3**, 45–70.
- Fagerlund B, Mackeprang T, Gade A, Hemmingsen R, Glenthøj BY (2004). Low-dose Zuclopenthixol on cognitive functions in first-episode drug-naïve schizophrenia patients. *CNS Spectrums* **9**, 364–374.
- Fatouros-Bergman H, Cervenka S, Flyckt L, Edman G, Farde L (2014). Meta-analysis of cognitive performance in drug-naïve patients with schizophrenia. *Schizophrenia Research* **158**, 156–162.
- Feldmann D, Schuepbach D, von Rickenbach B, Theodoridou A, Hell D (2006). Association between two distinct executive tasks in schizophrenia: a functional transcranial Doppler sonography study. *BMC Psychiatry* **6**, 25.
- Fioravanti M, Bianchi V, Cinti ME (2012). Cognitive deficits in schizophrenia: an updated metanalysis of the scientific evidence. *BMC Psychiatry* **12**, 64.
- Fioravanti M, Carlone O, Vitale B, Cinti ME, Clare L (2005). A meta-analysis of cognitive deficits in adults with a diagnosis of schizophrenia. *Neuropsychology Review* **15**, 73–95.
- Forbes N, Carrick L, McIntosh A, Lawrie S (2009). Working memory in schizophrenia: a meta-analysis. *Psychological Medicine* **39**, 889–905.
- Goel V (2002). Planning neural and psychological. In *Encyclopedia of Cognitive Science* (ed. L Nadel), pp. 697–703. Nature Publishing Group: London.
- Goldberg TE, Saint-Cyr JA, Weinberger R (1990). Assessment of procedural learning and problem solving in schizophrenic patients by Tower of Hanoi type tasks. *Journal of Neuropsychiatry and Clinical Neurosciences* **2**, 165–173.
- Greenwood KE, Wykes T, Sigmundsson T, Landau S, Morris RG (2011). Tower of London versus real life analogue planning in schizophrenia with disorganization and psychomotor poverty symptoms. *Journal of the International Neuropsychological Society* **17**, 474–484.
- Hedges L, Olkin I (1985). *Statistical Models for Meta-Analysis*. Academic Press: New York.
- Heinrichs RW, Zakzanis KK (1998). Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology* **12**, 426–445.
- Higgins JPT, Thompson SG (2004). Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine* **23**, 1663–1682.
- Hilti CC, Delko T, Orosz AT, Thomann K, Ludewig S, Geyer MA, Vollenweider FX, Feldon J, Cattapan-Ludewig K (2010). Sustained attention and planning deficits but intact attentional set-shifting in neuroleptic-naïve first-episode schizophrenia patients. *Neuropsychobiology* **61**, 79–86.
- Holt DV, Wolf J, Funke J, Weisbrod M, Kaiser S (2013). Planning impairments in schizophrenia: specificity, task independence and functional relevance. *Schizophrenia Research* **149**, 174–179.
- Huddy VC, Hodgson TL, Kapasi M, Mutsatsa SH, Harrison I, Barnes TRE, Joyce EM (2007). Gaze strategies during planning in first-episode psychosis. *Journal of Abnormal Psychology* **116**, 589–598.
- Humes GE, Welsh MC, Retzlaff P, Cookson N (1997). Towers of Hanoi and London: reliability and validity of two executive function tasks. *Assessment* **4**, 249–257.



- Joyce E, Hutton S, Mutsatsa S, Gibbins H, Webb E, Paul S, Robbins T, Barnes T (2002). Executive dysfunction in first-episode schizophrenia and relationship to duration of untreated psychosis: the West London Study. *British Journal of Psychiatry* **181**, 38–44.
- Jurado M, Rosselli M (2007). The elusive nature of executive functions: a review of our current understanding. *Neuropsychology Review* **17**, 213–233.
- Kafer KL, Hunter M (1997). On testing the face validity of planning/problem-solving tasks in a normal population. *Journal of the International Neuropsychological Society: JINS* **3**, 108–119.
- Kaller CP, Debelak R, Köstering L, Egle J, Rahm B, Wild PS, Blettner M, Beutel ME, Unterrainer JM (2016). Assessing planning ability across the adult life span: population-representative and age-adjusted reliability estimates for the Tower of London (TOL-F). *Archives of Clinical Neuropsychology* **31**, 148–164.
- Kaller CP, Rahm B, Köstering L, Unterrainer JM (2011a). Reviewing the impact of problem structure on planning: a software tool for analyzing tower tasks. *Behavioural Brain Research* **216**, 1–8.
- Kaller CP, Rahm B, Spreer J, Mader I, Unterrainer JM (2008). Thinking around the corner: the development of planning abilities. *Brain and Cognition* **67**, 360–370.
- Kaller CP, Rahm B, Spreer J, Weiller C, Unterrainer JM (2011b). Dissociable contributions of left and right dorsolateral prefrontal cortex in planning. *Cerebral Cortex* **21**, 307–317.
- Kaller CP, Unterrainer JM, Rahm B, Halsband U (2004). The impact of problem structure on planning: insights from the Tower of London task. *Cognitive Brain Research* **20**, 462–472.
- Kaller CP, Unterrainer JM, Stahl C (2012). Assessing planning ability with the Tower of London task: psychometric properties of a structurally balanced problem set. *Psychological Assessment* **24**, 46–53.
- Kay SR, Opler LA, Lindenmayer JP (1989). The Positive and Negative Syndrome Scale (PANSS): rationale and standardisation. *British Journal of Psychiatry. Supplement*, issue 7, 59–67.
- Knowles E, David A, Reichenberg A (2010). Processing speed deficits in schizophrenia: reexamining the evidence. *American Journal of Psychiatry* **167**, 828–835.
- Konstantopoulos S (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods* **2**, 61–76.
- Kontis D, Theochari E, Fryssira H, Kleisas S, Sofocleous C, Andreopoulou A, Kalogerakou S, Gazi A, Boniatsi L, Chaidemenos A, Tsaltas E (2013). COMT and MTHFR polymorphisms interaction on cognition in schizophrenia: an exploratory study. *Neuroscience Letters* **537**, 17–22.
- Köstering L, McKinlay A, Stahl C, Kaller CP (2012). Differential patterns of planning impairments in Parkinson's disease and sub-clinical signs of dementia? A latent-class model-based approach. *PLOS ONE* **7**, e38855.
- Köstering L, Nitschke K, Schumacher FK, Weiller C, Kaller CP (2015a). Test-retest reliability of the Tower of London planning task (TOL-F). *Psychological Assessment* **27**, 925–931.
- Köstering L, Schmidt CSM, Egger K, Amtage F, Peter J, Klöppel S, Beume L-A, Hoeren M, Weiller C, Kaller CP (2015b). Assessment of planning performance in clinical samples: reliability and validity of the Tower of London task (TOL-F). *Neuropsychologia* **75**, 646–655.
- Köstering L, Stahl C, Leonhart R, Weiller C, Kaller CP (2014). Development of planning abilities in normal aging: differential effects of specific cognitive demands. *Developmental Psychology* **50**, 293–303.
- Krabbendam L, de Vugt ME, Derix MM, Jolles J (1999). The behavioural assessment of the dysexecutive syndrome as a tool to assess executive functions in schizophrenia. *Clinical Neuropsychologist* **13**, 370–375.
- Langdon R, Coltheart M, Ward PB, Catts SV (2001). Mentalising, executive planning and disengagement in schizophrenia. *Cognitive Neuropsychiatry* **6**, 81–108.
- Langdon R, Coltheart M, Ward PB, Catts SV (2002). Disturbed communication in schizophrenia: the role of poor pragmatics and poor mind-reading. *Psychological Medicine* **32**, 1273–1284.
- Lezak MD, Howieson DB, Loring DW (2004). *Neuropsychological Assessment*. Oxford University Press: New York.
- Lowe C, Rabbitt P (1998). Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: theoretical and practical issues. Cambridge *Neuropsychological Test Automated Battery*. International study of post-operative cognitive dysfunction. *Neuropsychologia* **36**, 915–923.
- Marczewski P, de Linden MV, Larøi F (2001). Further investigation of the supervisory attentional system in schizophrenia: planning, inhibition, and rule abstraction. *Cognitive Neuropsychiatry* **6**, 175–192.
- McCormack T, Atance CM (2011). Planning in young children: a review and synthesis. *Developmental Review* **31**, 1–31.
- McKinlay A, Kaller CP, Grace RC, Dalrymple-Alford JC, Anderson TJ, Fink J, Roger D (2008). Planning in Parkinson's disease: a matter of problem structure? *Neuropsychologia* **46**, 384–389.
- Mesholam-Gately RI, Giuliano AJ, Goff KP, Faraone SV, Seidman LJ (2009). Neurocognition in first-episode schizophrenia: a meta-analytic review. *Neuropsychology* **23**, 315–336.
- Minzenberg MJ, Laird AR, Thelen S, Carter CS, Glahn DC (2009). Meta-analysis of 41 functional neuroimaging studies of executive function in schizophrenia. *Archives of General Psychiatry* **66**, 811–822.
- Miyake A, Friedman N, Emerson M, Witzki A, Howerter A, Wager T (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cognitive Psychology* **41**, 49–100.
- Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine* **6**, e1000097.
- Morice R, Delahunty A (1996). Frontal/executive impairments in schizophrenia. *Schizophrenia Bulletin* **22**, 125–137.
- Morris RG, Rushe T, Woodruffe PWR, Murray RM (1995). Problem solving in schizophrenia: a specific



- deficit in planning ability. *Schizophrenia Research* **14**, 235–246.
- Newman SD, Greco JA, Lee D (2009). An fMRI study of the Tower of London: a look at problem structure differences. *Brain Research* **1286**, 123–132.
- Newman SD, Pittman G (2007). The Tower of London: a study of the effect of problem structure on planning. *Journal of Clinical and Experimental Neuropsychology* **29**, 333–342.
- Nitschke K, Köstering L, Finkel L, Weiller C, Kaller CP (2017). A meta-analysis on the neural basis of planning: activation likelihood estimation of functional brain imaging results in the Tower of London task. *Human Brain Mapping* **38**, 396–413.
- Nitschke K, Ruh N, Kappler S, Stahl C, Kaller CP (2012). Dissociable stages of problem solving (I): temporal characteristics revealed by eye-movement analyses. *Brain and Cognition* **80**, 160–169.
- Norman D, Shallice T (1986). Attention to action: willed and automatic control of behavior. In *Consciousness and Self-Regulation: Advances in Research and Theory* (ed. R Davidson, G Schwartz and D Shapiro), pp. 1–18. Plenum: New York.
- Okruszek Ł, Rutkowska A (2013). Planning disorders in men with schizophrenia and in men with localized frontal lobe lesions. [Article in Polish] *Psychiatria Polska* **47**, 921–931.
- Owen A (2005). Cognitive planning in humans: new insights from the Tower of London (TOL) task. In *The Cognitive Psychology of Planning* (ed. R Morris and G Ward), pp. 135–151. Psychology Press: Hove.
- Palmer B, Dawes S, Heaton R (2009). What do we know about neuropsychological aspects of schizophrenia? *Neuropsychology Review* **19**, 365–384.
- Pantelis C, Barnes TRE, Nelson HE, Tanner S, Weatherley L, Owen AM, Robbins TW (1997). Frontal–striatal cognitive deficits in patients with chronic schizophrenia. *Brain* **120**, 1823–1843.
- R Core Team (2015). R: A Language and Environment for Statistical Computing (version 3.2.3) (<http://www.R-project.org>).
- Rainville C, Lepage E, Gauthier S, Kergoat M-J, Belleville S (2012). Executive function deficits in persons with mild cognitive impairment: a study with a Tower of London task. *Journal of Clinical and Experimental Neuropsychology* **34**, 306–324.
- Rohatgi A (2015). WebPlotDigitizer (version 3.9) (<http://arohatgi.info/WebPlotDigitizer>).
- Ruh N, Rahm B, Unterrainer JM, Weiller C, Kaller CP (2012). Dissociable stages of problem solving (II): first evidence for process-contingent temporal order of activation in dorsolateral prefrontal cortex. *Brain and Cognition* **80**, 170–176.
- Saleem MM, Harte MK, Marshall KM, Scally A, Brewin A, Neill JC (2013). First episode psychosis patients show impaired cognitive function – a study of a South Asian population in the UK. *Journal of Psychopharmacology* **27**, 366–373.
- Schaefer J, Giangrande E, Weinberger D, Dickinson D (2013). The global cognitive impairment in schizophrenia: consistent over decades and around the world. *Schizophrenia Research* **150**, 42–50.
- Schnirman GM, Welsh MC, Retzlaff PD (1998). Development of the Tower of London-Revised. *Assessment* **5**, 355–360.
- Shum D, Ungvari GS, Tang W-K, Leung JP (2004). Performance of schizophrenia patients on time-, event-, and activity-based prospective memory tasks. *Schizophrenia Bulletin* **30**, 693–701.
- Stanley TD, Doucouliagos H (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods* **5**, 60–78.
- Sullivan J, Riccio C, Castillo C (2009). Concurrent validity of the tower tasks as measures of executive function in adults: a meta-analysis. *Applied Neuropsychology* **16**, 62–75.
- Syvöaja HJ, Tammelin TH, Ahonen T, Räsänen P, Tolvanen A, Kankaanpää A, Kantomaa MT (2015). Internal consistency and stability of the CANTAB neuropsychological test battery in children. *Psychological Assessment* **27**, 698–709.
- Tenjin T, Miyamoto S, Miyake N, Ogino S, Kitajima R, Ojima K, Arai J, Teramoto H, Tsukahara S, Ito Y, Tadokoro M, Anai K, Funamoto Y, Kaneda Y, Sumiyoshi T, Yamaguchi N (2012). Effect of blonanserin on cognitive function in antipsychotic-naïve first-episode schizophrenia. *Human Psychopharmacology* **27**, 90–100.
- Thompson S, Higgins J (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine* **21**, 1559–1573.
- Tsuchimine S, Yasui-Furukori N, Kaneda A, Kaneko S (2013). Differential effects of the catechol-O-methyltransferase Val158Met genotype on the cognitive function of schizophrenia patients and healthy Japanese individuals. *PLOS ONE* **8**, e76763.
- Tunstall JR, O’Gorman JG, Shum DHK (2016). A four-disc version of the Tower of London for clinical use. *Journal of Neuropsychology* **10**, 116–129.
- Tyson PJ, Laws KR, Roberts KH, Mortimer AM (2004). Stability of set-shifting and planning abilities in patients with schizophrenia. *Psychiatry Research* **129**, 229–239.
- Unterrainer JM, Kaller CP, Loosli SV, Heinze K, Ruh N, Paschke-Müller M, Rauh R, Biscaldi M, Rahm B (2015a). Looking ahead from age 6 to 13: a deeper insight into the development of planning ability. *British Journal of Psychology* **106**, 46–67.
- Unterrainer JM, Owen AM (2006). Planning and problem solving: from neuropsychology to functional neuroimaging. *Journal of Physiology, Paris* **99**, 308–317.
- Unterrainer JM, Rahm B, Halsband U, Kaller CP (2005). What is in a name: comparing the Tower of London with one of its variants. *Cognitive Brain Research* **23**, 418–428.
- Unterrainer JM, Rauh R, Rahm B, Hardt J, Kaller CP, Klein C, Paschke-Müller M, Biscaldi M (2015b). Development of planning in children with high-functioning autism spectrum disorders and/or attention deficit/hyperactivity disorder. *Autism Research* **9**, 739–751.
- Unterrainer JM, Ruh N, Loosli SV, Heinze K, Rahm B, Kaller CP (2013). Planning steps forward in development: in girls earlier than in boys. *PLOS ONE* **8**, e80772.
- van Houwelingen H, Arends L, Stijnen T (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine* **21**, 589–624.

- Ventura J, Helleman GS, Thames AD, Koellner V, Nuechterlein KH** (2009). Symptoms as mediators of the relationship between neurocognition and functional outcome in schizophrenia: a meta-analysis. *Schizophrenia Research* **113**, 189–199.
- Viechtbauer W** (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* **36**, 1–48.
- Ward G, Allport A** (1997). Planning and problem-solving using the five-disc Tower of London task. *Quarterly Journal of Experimental Psychology* **50**, 49–78.
- Ward G, Morris R** (2005). Introduction to the psychology of planning. In *The Cognitive Psychology of Planning* (ed. G Ward and R Morris), pp. 1–34. Psychology Press: Hove.
- Zhu Y, Liu X, Wang H, Jiang T, Fang Y, Hu H, Wang G, Wang X, Liu Z, Zhang K** (2010). Reduced prefrontal activation during Tower of London in first-episode schizophrenia: a multi-channel near-infrared spectroscopy study. *Neuroscience Letters* **478**, 136–140.