Systematic Review and Meta-Analysis of Cognitive Interventions for Children With Central Nervous System Disorders and Neurodevelopmental Disorders

Kristen E. Robinson,1,2 PhD, Eloise Kaizar,3 PhD, Cathy Catroppa,4,5 PhD, Celia Godfrey,4 DPSYCH, and Keith Owen Yeates,1,2 PhD

1Department of Pediatrics, The Ohio State University, 2Center for Biobehavioral Health, The Research Institute at Nationwide Children’s Hospital, 3Department of Statistics, The Ohio State University, 4Clinical Sciences, Murdoch Childrens Research Institute, and 5Department of Psychology, Royal Children’s Hospital

All correspondence concerning this article should be addressed to Keith Owen Yeates, PhD, Department of Psychology, University of Calgary, 2500 University Drive NW, AD 254, Calgary, AB T2N 1N4, Canada. E-mail: kyeates@ucalgary.ca

Received January 31, 2014; revisions received April 10, 2014; accepted April 19, 2014

Objective To assess the efficacy of cognitive interventions for children with neurological disorders, acquired brain injuries, and neurodevelopmental disorders. Method We searched for randomized controlled trials of cognitive interventions; 13 studies met inclusion criteria. Risk of bias was rated for each study. Standardized effect size estimates were examined in 7 outcome domains. The overall quality of evidence was rated using the Grading of Recommendations Assessment, Development and Evaluation system. Results Significant positive treatment effects were found in all outcome domains aside from inhibitory control. Effects were large for attention, working memory, and memory tasks, and small for academic achievement and behavior rating scales. Results exhibited substantial heterogeneity in all domains. Overall quality of evidence was rated very low in all domains, suggesting substantial uncertainty about effect size estimates. Discussion The results provide some evidence of a positive benefit from cognitive interventions, but cannot be regarded as robust given the overall very low quality of the evidence.

Key words intervention outcome; meta-analysis; neurological disorders; randomized controlled trial; systematic review.

Pediatric neurological disorders (e.g., epilepsy, brain tumors), acquired brain injuries (e.g., traumatic brain injuries), and neurodevelopmental disorders (e.g., specific learning disabilities, autistic spectrum disorder; attention-deficit/hyperactivity disorder [ADHD]) all share a common feature, which is that they are frequently associated with deficits in domain-general cognitive functions such as attention, memory, and executive functions (Ozonoff, 2010; Ris & Abbey, 2010; Westerveld, 2010; Willcutt, 2010; Yeates, 2010). Deficits in these areas are thought to help account for the functional impairments that children with such disorders display across a range of settings, including academic difficulties, behavioral problems, and social isolation. A variety of interventions have been developed to ameliorate the cognitive deficits displayed by children with central nervous system and neurodevelopmental disorders, in hopes of improving their functional outcomes. Interventions for cognitive deficits typically involve a series of sessions, with the child either interacting directly with a therapist or, in computerized home-based programs, working under the supervision of a parent or adult caretaker. The sessions typically involve instruction and practice in specific cognitive tasks, with the goal of increasing the underlying skills of attention, memory, and/or executive functions. Therapist-delivered interventions may also include specific “homework” assignments that are completed at home, in some cases with the assistance of adult caretakers or parents. The interventions vary in
terms of treatment parameters, including timing (i.e., relative to onset of acute disorders), intensity (i.e., sessions per week; time spent per session), and duration (i.e., length of treatment program). The interventions also vary in their specificity, with some focusing on a single cognitive skill (e.g., working memory) and others taking a more comprehensive approach that encompasses multiple skills (e.g., metacognitive training in problem solving).

A number of systematic reviews have been published regarding cognitive interventions. Some have focused on children with neurodevelopmental disorders or acquired brain injuries (Laatsch et al., 2007; McCormick, Aubut, Gnanakumar, Curiale, & Marshall, 2012; Ross, Dorris, & McMillan, 2011; Slomine & Locascio, 2009; Wolfe, Madan-Swain, & Kana, 2012) and others on children with neurodevelopmental disorders (Rapport, Orban, Kofler, & Friedman, 2013; Toplak, Connors, Shuster, Knezovic, & Parks, 2008). Some have focused on specific interventions (e.g., working memory training; Melby-Lervåg & Hulme, 2013), whereas others are more general. Most of the reviews have classified studies in terms of evidence quality, but few have involved quantitative meta-analysis. The findings from the reviews can be broadly characterized as mixed. In most cases, the reviews find evidence of short-term improvements in performance on specific cognitive tasks, but limited evidence of sustained benefits over time. Evidence of generalization of cognitive interventions to broader forms of academic, behavioral, and social functioning is less convincing than the evidence for cognitive gains.

The current review was intended to determine the efficacy of cognitive interventions for children with neurodevelopmental disorders, acquired brain injuries, and neurodevelopmental disorders. The review was limited to randomized controlled trials (RCT) and other controlled clinical trials comparing cognitive interventions with an attention-only control, other active treatment, or waiting list control. Studies were eligible for inclusion only if participants were <19 years old and they had a neurological disorder, acquired brain injury, or neurodevelopmental disorder. We did not include studies of healthy children or those identified solely as intellectually disabled. Interventions needed to be primarily cognitive in nature (e.g., we excluded interventions focused on social communication in autistic spectrum disorders) and have credible recognizable cognitive content. Cognitive interventions were defined as any treatment specifically designed with the intention of improving child outcomes in attention, memory, or executive functions (e.g., working memory, inhibitory control).

The review adds to the existing literature in several regards. Perhaps most importantly, no previous review has examined the efficacy of cognitive interventions across neurodevelopmental disorders, neurological disorders, and acquired brain injuries. Given that the target of the interventions is cognitive functioning, rather than a specific diagnosis, we expected that pooling studies across disorders would provide a more robust estimate of treatment efficacy and enable a quantitative meta-analysis. The review also incorporates newer studies that do not appear in previous reviews. Additionally, the review focuses specifically on attention, memory, and executive functions, cognitive domains often considered as critical underpinnings of functional outcomes.

Method
Selection of Studies
A comprehensive search of existing peer-reviewed studies was conducted using multiple strategies. First, keyword database inquiries were conducted. These included searches in PubMed, PsycInfo, Embase, and the Cochrane Central Register of Controlled Trials (CENTRAL). In each search, keywords were grouped as either diagnosis terms (e.g., brain injuries, learning disability) or intervention terms (e.g., treatment outcome, rehabilitation), and all possible combinations of keywords were used. An example search strategy including all search terms can be found in Supplementary Table I. Database filters limited the search to articles published between 1980 and May 2013. Second, reference lists from articles found in the initial search were reviewed and additional potential studies were identified. Third, relevant review articles and meta-analyses were similarly reviewed and their reference lists were also examined. After eliminating duplicate articles, a total of 623 studies published between 1980 and 2013 were identified and reviewed for potential inclusion in the meta-analysis.

Requirements for Inclusion
To be eligible for inclusion in the systematic review, each study was required to meet all of the following inclusion criteria. First, studies must have included a sample of children or adolescents who were diagnosed with either a neurodevelopmental (e.g., ADHD, specific learning disability) or central nervous system (e.g., epilepsy, traumatic brain injury, brain tumor) disorder. We did not require that a study specifically state that participants met diagnostic criteria (e.g., Diagnostic and Statistical Manual of Mental Disorders, 4th Edition criteria) although most included studies did so. Second, studies must have included only participants aged <19 years; if both older and younger participants were included, data on child and adolescent participants must have been presented separately.
Third, study designs were limited to RCT and other controlled clinical trials that included at least two points of measurement, one at baseline and another at a point in time afterward to measure the efficacy of the intervention. Fourth, studies must have involved an intervention with a primary aim of promoting attention, memory, or executive function, including metacognition. Fifth, the treatment arm and all control/comparison arms must have included at least 10 participants at the end of treatment or follow-up. Finally, studies must have been published in English. We excluded studies that focused on speech/language/communication skills or specific academic skills (e.g., reading) because they are not typically considered cognitive interventions. We also excluded studies that targeted only parents.

Primary outcomes of interest were tests of specific cognitive skills (both standardized and experimental). Secondary outcomes included rating scales meant to measure behaviors that reflect the everyday manifestation of specific cognitive skills (e.g., Behavior Rating Inventory of Executive Function for executive functions) and other functional domains thought to be promoted by the intervention (e.g., academic skills, behavioral adjustment).

**Determination of Eligibility**

The process of study selection and determination of eligibility is summarized in graphical form consistent with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines in Figure 1. Abstracts from each of 623 original studies were reviewed for potential inclusion in the meta-analysis. Each abstract was randomly assigned for review using a random number generator. Abstracts were each reviewed by two independent reviewers and were categorized as (a) likely eligible, (b) maybe eligible, or (c) not eligible. Inter-reviewer agreement was perfect for 549 of the 623 articles (88.1%). All articles categorized by at least one of the reviewers as likely or maybe eligible underwent a full-text review by one reviewer, using a standard data collection form. Full-text reviews were completed for 117 articles, which were again randomly assigned for review using a random number generator. Of the 117 articles that underwent full-text review, 104 did not meet all of the inclusion criteria for the systematic review. Studies were excluded for a variety of reasons, the most common being that they did not involve a randomized or controlled clinical trial, lacked a control or comparison arm, had <10 participants at the follow-up assessment, or did not report on an intervention targeting attention, memory, or executive function (see Figure 1). Several studies were excluded because their participants had a disorder that could affect brain function, but only as a secondary effect rather than a primary consequence of the disease or injury (e.g., low birth weight, HIV); we excluded those studies because brain impairment is not characteristic of all children with those disorders.

The full-text coding process yielded 13 original studies that met inclusion criteria for the meta-analysis (see studies preceded by asterisk in References). The data presented in some papers were not adequate to support meta-analysis; we contacted the authors of those papers and asked that they provide whatever data were necessary for inclusion. Sufficient data for inclusion were available for 12 studies; one study (Chan & Fong, 2011) was excluded from the meta-analysis but was included in qualitative reviews. Table I provides descriptive information about each of the 13 included studies.

**Assessment of Risk of Bias**

The methodological quality of the 13 studies included in the systematic review was rated using the Cochrane Risk Bias Tool (Higgins & Green, 2011). This procedure involves rating each study for several potential sources of bias: selection bias (i.e., random sequence generation for assignment to treatment conditions and allocation concealment); performance bias (i.e., blinding of participants and study personnel); detection bias (i.e., blinding of outcome assessment); attrition bias (i.e., incomplete outcome data); and reporting bias (i.e., selective reporting of results). Each form of bias is rated low, high, or unclear based on the information provided in the published article. Other forms of bias may also be rated if noted by the reviewer.

**Data Analysis**

To facilitate data analysis, we first identified seven major categories of outcomes for which at least two studies presented data, as listed in Table II. The table also lists the number of studies that included outcome measures within each outcome category, and the total number of outcome measurements used across those studies. Most studies included working memory tasks (8 of 13 studies) and behavioral rating scales meant to assess attention (8 of 13 studies), but fewer than half of the studies addressed any of the other categories. Unfortunately, the small number of studies and measurements within each category was exacerbated by the diverse choices of outcome metrics. Within each category, most studies chose unique measures. The most consistent choice of metric was for behavioral rating scales measuring attention, where five studies used the Connors ADHD parent rating scale (Conners, 1997) and three studies used the Connors Cognitive Problems and Hyperactivity parent rating scales (Conners, 1997). Four studies used a digit span task to measure working
memory. No other measure was used in more than two studies.

To combine effects across studies to estimate an overall effect size, we first defined a standardized effect size to be the difference between treatment groups divided by the population standard deviation:

$$\delta = \frac{\mu_T - \mu_C}{\sigma} = \frac{(\mu_{T1} - \mu_{T0}) - (\mu_{C1} - \mu_{C0})}{\sigma}$$

where T indicates the active or more intense treatment group, C indicates the control or less intense treatment group, 1 indicates the posttreatment time point, 0 indicates the pretreatment time point, and \(\sigma\) indicates the standard deviation within treatment and period. A positive effect indicates that the active or more intense treatment was beneficial compared with the control or less intense treatment group. We follow the standard repeated measures analysis of variance assumption that the standard deviation is constant across the groups and time points.

Further, because most authors only report pre- and posttreatment measures, we disregarded any follow-up measurements.

Most authors did not report the standardized difference and its standard error. However, most authors did publish or provide us with enough information to calculate reasonable estimates of the effect of interest, \(\delta\). Supplementary Table II summarizes how each reported statistic or collection of statistics was transformed to provide an estimate and the standard error. However, some studies provided only group-by-time-specific means and measures of variability. In that case, we could not exploit the pre–post nature of the study to correctly estimate the standard deviation of the change in outcome across participants. We conservatively assumed that the pre- and postmeasures for each individual are independent, thus producing a much larger standard error for the effect than would most likely have been calculated with complete

---

**Figure 1. PRISMA flowchart.**
| Study | Target group | Comparison group | Randomization procedure | Intervention type | Intervention characteristics | Outcomes assessed |
|-------|--------------|------------------|------------------------|------------------|-----------------------------|-------------------|
| Beck, Hanson, | ADHD | ADHD | Alternating assignment in alphabetical order | Cogmed computerized working memory training | | Attention, working memory, metacognition |
| Pufkenberger, Benninger, & Benninger, 2010 | | | | | | |
| Braga, Rossi, Moretto, da Silva, & Cole, 2012 | ABI | ABI | Metacognitive dimension program | | | Working memory, metacognition |
| Butler et al., 2008 | Malig | Malig | Randomized by "central data manager" in 2:1 ratio | Cognitive remediation program; brain injury rehabilitation, educational, and child clinical psychology approaches | | Attention, memory, working memory, academic achievement, reaction time |

(continued)
| Study | Target group | Comparison group | Randomization procedure | Intervention type | Intervention characteristics | Outcomes assessed |
|-------|--------------|------------------|-------------------------|------------------|----------------------------|-------------------|
| Chan & Fong, 2011 | TBI, Malig, AVM | TBI, Malig, AVM | Coin toss to randomly assign matched pairs | Problem-solving skills training program | Center based (group) | Metacognition, general executive function |
| Gibson et al., 2011<sup>a</sup> | ADHD | ADHD | Randomized by a Cogmed programmer in consecutive fashion by allocation sequence ID number | Cogmed-RM program split into verbal working memory and spatial working memory groups | Home based | Attention, working memory |
| Gray et al., 2012 | LD and ADHD | LD and ADHD | Randomized by a senior researcher using a random numbers table, separately for each cohort, stratified for gender, in a 3:2 ratio | Cogmed computerized working memory training | School based | Attention, working memory, academic achievement |

<sup>a</sup> ADHD 24 19 12.33 (1.17) ADHD 23 18 12.87 (1.25) Randomized by a senior researcher using a random numbers table, separately for each cohort, stratified for gender, in a 3:2 ratio.
| Study                          | Target group | Comparison group |
|-------------------------------|--------------|------------------|
| Sample                        | Number       | Number           |
| M (SD)                        | M (SD)       |                  |
| Green et al., 2012            | ADHD 12      | ADHD 14          |
|                               | 8            | 9                |
|                               | 9.9 (1.8)    | 9.6 (2.6)        |
| Randomization procedure       | Randomized   | Cogmed computer- |
|                               | using a      | 25 sessions      |
|                               | computer-    | 40 min each      |
|                               | generated    | 25-day tx window |
|                               | random       | Parents coached  |
|                               | number       | by licensed      |
|                               | sequence     | clinical         |
|                               |              | psychologist     |
| Johnstone, Roodenrys, Phillips, Watt, & Mantz, 2010 | ADHD 15 | ADHD 14 |
|                               | 13           | 12               |
|                               | 11.0 (0.4)   | 10.8 (0.4)       |
| Random allocation completed by | Random       | Computerized     |
|                               | one researcher| training program | Home based |
|                               | (S.R.)       | with working     |
|                               |              | memory and       |
|                               |              | response inhibition games |
|                               |              | Home based |
|                               |              | 25 sessions |
|                               |              | 20 min each |
|                               |              | 5-week tx window |
|                               |              | No information |
|                               |              | on parent |
|                               |              | supervision |
| Johnstone et al., 2012        | ADHD 40      | ADHD 20          |
|                               | 35           | 19               |
|                               | 9.7 (2.1)    | 9.9 (2.3)        |
| Randomization procedure       | ND           | Computerized     |
|                               |              | training program |
|                               |              | with working     |
|                               |              | memory and       |
|                               |              | response inhibition games |
|                               |              | Home based |
|                               |              | 25 sessions |
|                               |              | 20 min each |
|                               |              | 5-week tx window |
|                               |              | No apparent |
|                               |              | supervision of |
|                               |              | parents |
| Klingberg et al., 2005        | ADHD 20      | ADHD 24          |
|                               | 16           | 20               |
|                               | 9.8 (1.4)    | 9.7 (1.3)        |
| Randomization procedure       | Randomized   | Cogmed computer- |
|                               | using a      | 25 sessions      |
|                               | blinded list | 40 min each      |
|                               | of ID numbers| 25-day tx window |
|                               | associated   | Parents coached  |
|                               | with        | by licensed      |
|                               | software     | clinical         |
|                               | CD's        | psychologist     |
|                               | distributed  |                 |
|                               | in         |                 |
|                               | consecutive |                 |
|                               | order       |                 |
|                               |              | Attention, working memory, hyperactivity |
|                               |              | Inhibitory control |
|                               |              | Attention, working memory, inhibitory control |

(continued)
Table I. Continued

| Study                          | Target group | Comparison group |
|-------------------------------|--------------|-----------------|
|                               | Sample | N | Number of male | Age M (SD) | Sample | N | Number of male | Age M (SD) | Randomization procedure | Intervention type | Intervention characteristics | Outcomes assessed |
| Shalev, Tsal, & Mevorach, 2007 | ADHD 20 | 17 | 9.1 | ADHD 16 | 13 | 9.2 | ND | Computerized progressive attentional training program | Center based | 16 sessions 1 hr each 8-week tx window Treatment provided by research assistants | Attention, academic achievement |
| Solan, Shelley-Tremblay, Ficarra, Silverman, & Larson, 2003 | LD 15 | 11.3 (0.3) | LD 15 | 11.3 (0.3) | ND | Attention enhancing programs | School based | 12 sessions 1 hr each 12-week tx window Treatment provided by the investigators | Attention, academic achievement |
| Van’t Hooft et al., 2007 | TBI Infection Malig 18 | 12 | 13.50 (2.51) | TBI Infection Malig 20 | 10 | 14.65 (2.86) | Consecutive random assignment | Amsterdam memory and attention training for children (Amat-c) program | Home or school based | 102 sessions 30 min each 17-week tx window Teacher or parent coach oversee individual sessions; weekly on-site supervision by research staff | Attention, working memory, memory, inhibitory control |

Note. ADHD = attention deficit/hyperactivity disorder; ABI = unspecified acquired brain injury; TBI = traumatic brain injury; Malig = Malignancy with CNS involvement (brain tumor or leukemia with CNS involvement); AVM = arteriovenous malformation; LD = learning disabilities; Infection = e.g., encephalitis; M = Mean; SD = standard deviation; ND = no data provided. For studies other than those indicated below, data presented in the “Comparison Group” column describe participants assigned to a waitlist control/treatment as usual condition. 

1Gibson et al., 2011, involved a comparison of two types of intervention. Data presented in the “Intervention Group” column describe characteristics of the verbal working memory group and data presented in the “Comparison Group” column describe characteristics of the spatial working memory group. Intervention details and outcome measures apply to both groups.

1Johnstone et al., 2010, involved a comparison of two intensities of intervention. Data presented in the “Intervention Group” column describe characteristics of the high-intensity intervention group and data presented in the “Comparison Group” column describe characteristics of the low intensity intervention group.
Table II. Number of Studies and Outcome Measurements by Outcome Domain

| Outcome domain                        | Studies | Total number of measurements |
|---------------------------------------|---------|------------------------------|
| Attention tasks                       | 4       | 14                           |
| Working memory tasks                  | 7       | 19                           |
| Memory tasks                          | 2       | 11                           |
| Inhibitory control tasks              | 3       | 3                            |
| Attention Behavior Rating Scales      | 8       | 26                           |
| Working Memory Behavior Rating Scales | 3       | 4                            |
| Academic achievement tests            | 4       | 14                           |

Although many of the included studies used a large number of outcome measures, none of the published data provided information about correlation among effect sizes that would allow us to correctly use more than one measure per study in any estimation of effect size within domain. Thus, we took two separate approaches to the estimation of an overall effect size. First, we took a traditional meta-analytic approach where we used the inverse variance estimator under the unrealistic assumption that the effect on each measure is independent across all the studies (DerSimonian & Laird, 1986). We labeled this the “Overall Mean Effect.” We next relaxed this assumption to fit a weighted linear random-effects model that accommodates within-study correlation. We called this the “Hierarchical Mean Effect.”

The protocol for the review initially sought to determine the relative efficacy of interventions along six major dimensions: type of participants (i.e., are interventions more effective for neurological disorders or acquired brain injuries versus neurodevelopmental disorders?); age of participants (i.e., are interventions more effective for older than younger children?); parent involvement in intervention (i.e., are interventions more effective when they include a parent component than when they focus solely on children?); methodological quality (i.e., is efficacy a function of methodological quality of the selected studies?); transfer effects (i.e., is efficacy a function of whether outcomes are similar to those that have been trained?); and early versus late outcomes (i.e., is efficacy sustained over time after the completion of the intervention?). Unfortunately, the small number of studies and nature of the data available in those studies precluded most moderator analyses. We were able to assess differences in effect size related to type of participant (i.e., neurological disorder/acquired brain injury versus neurodevelopmental disorder) in two domains (i.e., attention tasks and working memory tasks), which were the only ones that included at least two studies of each type of participant. In each case, we used a weighted random-effects hierarchical model with a fixed effect for type of participant. We used the Kenward-Roger adjusted F-test to test for an effect of participant type (Kenward & Roger, 1997).

Heterogeneity of effects was assessed based on the $I^2$ metric, which quantifies the proportion of an observation’s total variability that is due to study-to-study variation (Higgins & Thompson, 2002). We used the $I^2$ metric directly to assess heterogeneity for outcome categories with only one measurement per study (i.e., inhibitory control). For the other outcome categories, we calculated an approximate $I^2$ value by treating the study-level random effect estimate as if it were known in a standard random-effects meta-analysis. Treating the study-specific estimate as if it were known in a standard random-effects meta-analysis, we believe this overestimation to be small, since similar calculations based on hierarchical models result in similar estimates of heterogeneity.

Finally, we were concerned about publication bias because many studies openly noted that they reported only those results that were deemed “significant” (i.e., generally with $p < .05$). For these studies, we could not include any estimate for those effects that were not “significant”; even though we could estimate the effect size to be zero, we did not have a corresponding standard error necessary for including the effect in the meta-analysis. Because Johnstone et al. (2012) explicitly noted nonsignificant tests for relevant measures, but did not report corresponding statistics that could be converted into an effect estimate, we note an “NA” for these measures in the forest plots. In addition to incomplete results, entire studies may exist that were not reported in the literature or that we did not discover in our search. To address these concerns, we examined funnel plots and used the “trim and fill” method to approximate an overall effect size when a funnel plot showed evidence of publication bias. For this rough approximation, we used the estimated within-study overall effect size (“Study Mean Effect”) as a single observation, and “filled in” potentially missed studies to balance the funnel plot. We take some of these “filled-in” studies to represent the average “nonsignificant” effects for the measures that had not been reported. Because this method is based entirely on the
supposed symmetry of our incomplete study-specific effect sizes, we do not interpret these as true estimates, but rather as indications of the direction and magnitude of possible publication bias.

All calculations were carried out using R software version 3.0.1, including the meta package version 3.0.1 and the lme4 package version 1.0.5.

Assessment of Quality of Evidence
To complement the meta-analytic results, we also assessed the overall quality of the evidence for each outcome using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system (Guyatt et al., 2011). In the GRADE system, quality of evidence is rated very low, low, moderate, or high across all studies within a particular outcome domain. Randomized trials are initially considered to provide high-quality evidence. Five factors can lead to rating the quality of evidence lower when considered across studies (i.e., risk of bias; inconsistency in results across trials; indirectness of outcome measurement; imprecision in effect size estimates resulting from small samples or few studies; and publication bias). Three factors can lead to higher quality ratings (i.e., large magnitude of effect; dose–response gradient; and confounders minimize effect).

Results
Summary of Included Studies
The 13 included studies, published from 2003 to 2012, included 643 children in trials examining the efficacy of cognitive interventions. Three studies involved children with central nervous system disorders, most often acquired brain injuries or brain tumors; the other nine studies involved children with neurodevelopmental disorders, with seven of those focusing exclusively on ADHD. Most of the interventions involved computerized training, often CogMed. The interventions were not typically delivered by psychologists, but instead were supervised by parents in children’s homes. The number of sessions ranged from 12 to 102 (median = 25), with sessions ranging in length from 20 min to 2–3 hr (median = 45 min). The duration of the interventions ranged from 25 days to 20 weeks (median = 6 weeks). Treatment fidelity was reportedly assessed in nine studies, most often through reviews of computerized session data or session videotapes; however, four studies provided no information regarding treatment fidelity checks.

Risk of Bias
Figure 2 provides a summary of the risk of bias ratings across the 13 included studies; ratings for each individual study and their justification are available in Supplementary Figure 1. The most significant source of bias was blinding of participants and study personnel, where 50% of the studies were judged to have a high risk of bias. Blinding of study participants and study personnel can be difficult in behavioral interventions, and so this finding is not surprising. A high risk of bias in the blinding of outcome assessment also was fairly common, occurring in 27% of the studies. The risks of bias in allocation concealment (i.e., where investigators could foresee assignment to conditions) and in reporting of outcome data were considered unclear in many studies, largely because of incomplete reporting in the published studies.

Effects of Intervention
Figures 3–9 provide forest plots representing the results of the meta-analysis within each outcome domain. We considered overall mean effects of 0.20 to be small, 0.50 to be medium, and 0.80 to be large (Cohen, 1988).

Attention Tasks
Four studies, which used 14 different measurements, examined the effects of cognitive interventions on attention tasks. Both the overall mean effect (0.860, 95% confidence interval [CI]: 0.410, 1.309) and the hierarchical mean effect (0.320, 95% CI: 0.241, 0.402) were positive and significant (i.e., 95% CIs did not include 0), with the former being large in magnitude and the latter being small to medium. The hierarchical mean effect was substantially reduced relative to the overall mean effect for attention tasks (i.e., to a small-to-medium effect) because one study (Van’t Hooft et al., 2007) provided the majority of the positive effect estimates.

Working Memory Tasks
Intervention effects on working memory tasks were examined in seven studies that used 19 different measurements.
Both the overall mean effect (0.756, 95% CI: 0.498, 1.104) and the hierarchical mean effect (0.799, 95% CI: 0.712, 0.885) were positive, significant, and large in magnitude.

Memory Tasks
Only two studies examined intervention effects on memory tasks, using 11 different measurements. Both the overall mean effect (0.953, 95% CI: 0.619, 1.287) and the hierarchical mean effect (1.235, 95% CI: 1.035, 1.436) were positive, significant, and large in magnitude.

Inhibitory Control Tasks
Three studies, which used three different measurements, examined intervention effects on inhibitory control tasks. Neither the overall mean effect (0.594, 95% CI: −0.025, 1.212) nor the hierarchical mean effect (0.541, 95% CI: −0.032, 1.114) was significant, although they were both positive and medium in magnitude.

Attention Behavior Rating Scales
Eight studies examined intervention effects on attention behavior rating scales, using 26 different measurements. Both the overall mean effect (0.296, 95% CI: 0.205, 0.387) and the hierarchical mean effect (0.321, 95% CI: 0.293, 0.350) were positive, significant, and small to medium in magnitude.
Figure 4. Forest plot of effect size estimates for working memory tasks.
Working Memory Behavior Rating Scales

Intervention effects on working memory behavior rating scales were examined in three studies that used four different measurements. Both the overall mean effect (0.227, 95% CI: 0.054, 0.401) and the hierarchical mean effect (0.221, 95% CI: 0.194, 0.248) were positive, significant, and small in magnitude.

Academic Achievement Tests

Four studies, using 14 different measurements, examined intervention effects on academic achievement tests. Both the overall mean effect (0.145, 95% CI: 0.066, 0.223) and the hierarchical mean effect (0.353, 95% CI: 0.314, 0.391) were positive and significant; the former was small in magnitude and the latter was small to medium.
Figure 7. Forest plot of effect size estimates for attention behavior rating scales.
Figure 8. Forest plot of effect size estimates for working memory behavior rating scales.

Figure 9. Forest plot of effect size estimates for academic achievement tests.


**Participant Group Differences**

Analyses comparing the effects of interventions for children with neurological disorders or acquired brain injuries versus those with neurodevelopmental disorders revealed significantly larger average effect sizes for the former group. For attention tasks, the average effect size for neurological disorders and acquired brain injuries was 0.314 ($F = 7.50, p = .006$) larger than for neurodevelopmental disorders. For working memory tasks, the average effect size for acquired brain injuries was 1.050 ($F = 99.90, p < .001$) larger than for ADHD.

**Heterogeneity of Effects**

Table III summarizes the heterogeneity of effects within outcome domains. According to the standards suggested in the Cochrane Handbook (Higgins & Green, 2011), the results within all domains exhibit substantial (i.e., $I^2$ between 50 and 90%) or considerable (i.e., $I^2$ greater than 75%) heterogeneity. The only exception was for behavior rating scales for working memory. This exception is not surprising, as the standard DerSimonian and Laird (1986) method underestimates the between-study variance to be zero and thus forces $I^2$ to be also zero. In general, the heterogeneity results confirm our decision to use a random-effects meta-analysis approach.

**Publication Bias**

Funnel plots provided evidence of significant potential publication bias in many outcome domains. After adding potentially missing studies using trim-and-fill methods, the estimated effect size was reduced substantially for attention tasks, working memory tasks, and academic achievement. Publication bias was not substantial for memory or inhibitory control tasks or for behavior rating scales measuring attention or working memory.

**Quality of Evidence**

Table IV provides a summary of the findings within each domain. In all cases, the quality of evidence was judged to be very low based on the GRADE criteria. Thus, we are uncertain about the effect size estimates in all domains. Study limitations were considered serious in all outcome domains based on the risk of bias judgments. Results were judged inconsistent in all domains other than behavior rating scales for working memory based on the heterogeneity of effects. Results were judged indirect for behavior rating scales, which do not directly assess cognitive or academic abilities. Imprecision was considered serious for studies that examined the domains of memory, inhibitory control, and behavior ratings of working memory, all of which had three or fewer studies and <200 total participants. Publication bias was considered serious or very serious for studies of attention tasks, working memory tasks, and academic achievement. Studies of attention tasks and working memory tasks were credited for large effect sizes.

**Discussion**

The results of the systematic review can be viewed as a glass half empty or half full. On one hand, the meta-analyses provide evidence of the benefits of cognitive interventions, with significant and positive mean effects in all outcome domains aside from inhibitory control. They were large in magnitude (i.e., >0.75) for attention, working memory, and memory tasks, and small in magnitude (i.e., <0.3) for academic achievement and for behavior rating scales assessing attention and working memory. This finding is consistent with the results of other meta-analyses, suggesting that cognitive interventions promote greater change in the cognitive skills they target than in broader measures of function (Melby-Lervåg & Hulme, 2013). Nevertheless, significant improvement was demonstrated in all domains. This is largely consistent with the conclusions of other previous reviews.

On the other hand, the overall quality of evidence was judged to be very low in all outcome domains based on the GRADE ratings, so that we are uncertain about the estimates of treatment efficacy. As is often the case for behavioral interventions, most studies suffered from some risk of bias, usually because of the lack of blinding of participants or study personnel. Another significant issue was the inconsistency of outcome measures, which we believe greatly contributed to the substantial heterogeneity in results within outcome domains. The lack of standards or guidelines in this regard has been problematic. The limited number of studies and participants further reduced the precision of effect size estimates. Publication bias was likewise substantial, leading to major reductions in the estimated effects of intervention on several outcomes. All in
Table IV. Summary of Findings by Outcome Domain

| Outcome domain                      | Mean (10th and 90th percentiles) of estimated effects in the control groups [number of control group measurements] | Change in effect with intervention (95% confidence limits)a | Number of participants (number of studies) | Quality of evidenceb | Comments |
|-------------------------------------|---------------------------------------------------------------------------------------------------------------|----------------------------------------------------------|---------------------------------------------|----------------------|----------|
| Attention tasks                     | −0.59 (−7.96, 6.57) [14]                                                                                      | 0.32 (0.24, 0.40)                                          | 270 (4)                                     | ⬤⬤⬤⬤               | –        |
| Working memory tasks                | −0.76 (−2.90, 0.41) [13]                                                                                      | 0.86 (0.77, 0.95)                                          | 386 (7)                                     | ⬤⬤⬤⬤               | Johnstone et al. (2012): Results only presented for significant comparisons; Gibson et al. (2011): No clear control/treatment designation; Gibson et al. (2011): ‘‘correct recall’’ only reported ‘‘not significant’’ |
| Memory tasks                        | 0.67 (−3.10, 0.50) [11]                                                                                       | 1.24 (1.03, 1.44)                                          | 180 (2)                                     | ⬤⬤⬤⬤               | –        |
| Inhibitory control tasks            | 0.46 (0.03, 0.89) [2]                                                                                         | 0.54 (−0.52, 1.60)                                         | 127 (3)                                     | ⬤⬤⬤⬤               | Johnstone et al. (2012): Flanker task % correct—incongruent: Nonsignificant effect size not reported |
| Attention Behavior Rating Scales    | −0.18 (−0.99, 0.57) [24]                                                                                      | 0.34 (0.31, 0.36)                                          | 396 (7)                                     | ⬤⬤⬤⬤               | Gibson et al. (2011): No clear control/treatment designation; Gray et al. (2012): SWAN inattention/hyperactive (parent) rounded from 0; Gibson et al. (2011): DuPaul Inattention (teacher) direction of nonsignificant effect not reported; Gibson et al. (2011): DuPaul Hyperactivity (parent): Greater improvement for visual intervention; Gibson et al. (2011): DuPaul ADHD Rating Scale (parent): Direction of nonsignificant effect not reported |
| Working Memory Behavior Rating Scales| 0.32 (0.15, 0.45) [3]                                                                                         | 0.24 (0.21, 0.27)                                          | 98 (2)                                      | ⬤⬤⬤⬤               | –        |
| Academic achievement                | 0.30 (−0.92, 1.56) [11]                                                                                       | 0.35 (0.31, 0.39)                                          | 253 (4)                                     | ⬤⬤⬤⬤               | –        |

Note. aCalculation of relative effects was not possible with the available data.

GRADE Working Group grades of evidence:
- ⬤⬤⬤⬤: High quality: Further research is very unlikely to change our confidence in the estimate of effect.
- ⬤⬤⬤⬤: Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
- ⬤⬤⬤: Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
- ⬤⬤: Very low quality: We are very uncertain about the estimate.
all, the limited quality of the evidence provides little confidence in the robustness of the meta-analytic results.

Because of the small number of studies and total participants within any given outcome domain, we were severely limited in our ability to conduct the moderator analyses originally planned. However, we did find evidence for larger effects in children with neurological disorders or acquired brain injuries than in children with neurodevelopmental disorders, at least for attention tasks and working memory tasks, which were the only two outcome domains that included at least two studies of each type of disorder. The reasons for larger effects in children with neurological disorders and brain injuries are not clear. Moreover, given the limited number of studies involved and the limited quality of the evidence overall, this finding must be considered cautiously, and certainly warrants replication.

Thus, the results of the systematic review are far from definitive. They provide some evidence of a positive benefit from cognitive interventions, especially for children with neurological disorders and acquired brain injuries, but the findings cannot be regarded as robust or reliable based on the overall very low quality of the evidence. Thus, the findings do not form the basis for strong recommendations regarding the use of cognitive interventions in clinical practice, although they suggest sufficient potential benefit to warrant further study. In this regard, the review highlights the need for additional clinical trials that evaluate the efficacy, as well as effectiveness, of cognitive interventions. Controlled trials comparing different interventions to one another and to appropriate controls may be especially informative.

To add appreciably to the existing literature, future studies need to be methodologically rigorous, preferably meeting the Consolidated Standards of Reporting Trials (CONSORT) guidelines for clinical trials (Schulz, Altman, Moher, & CONSORT Group, 2010). They should be designed to measure longer-term effects, rather than only immediate posttreatment outcomes, to determine whether any benefits of the interventions are sustained. Further, they should strive to use widely accepted outcome measures to enable cross-study comparisons, such as those recommended by the NIH Common Data Elements Project (e.g., McCauley et al., 2012). Lastly, to facilitate future systematic reviews, publications based on future trials should provide complete statistical data for all outcome measures, ideally to include standardized differences and their standard errors, as well as the correlations among effects when more than one outcome is assessed.

We acknowledge that conducting rigorous clinical trials is a significant challenge, and that the exigencies of journal word limits and editorial constraints sometimes prevent the full publication of all study details and results. Moreover, the current fiscal environment makes it difficult for investigators to obtain funding to support the development of promising interventions or to conduct large-scale clinical trials evaluating the efficacy of existing interventions. Only about half of the studies included in the review reported support through external or internal funding. Nevertheless, cognitive interventions are typically fairly intensive, and require substantial investment in time and energy both by children and their parents. Moreover, some cognitive interventions are being commercially marketed to families and children, at substantial cost to them not only in time but also financially. This raises concerns about the costs versus benefits of cognitive interventions, especially given their uncertain efficacy. Thus, we believe additional rigorous research is imperative.

Conflicts of interest: None declared.

References

Beck, S. J., Hanson, C. A., Puffenberger, S. S., Benninger, K. L., & Benninger, W. B. (2010). A controlled trial of working memory training for children and adolescents with ADHD. *Journal of Clinical Child and Adolescent Psychology, 39*, 825–836.

Braga, L. W., Rossi, L., Moretto, A. L. L., da Silva, J. M., & Cole, M. (2012). Empowering preadolescents with ABI through metacognition: Preliminary results of a randomized clinical trial. *NeuroRehabilitation, 30*, 205–212.

Butler, R. W., Copeland, D. R., Fairclough, D. L., Mulhern, R. K., Katz, E. R., Kazak, A. E., ... Sahler, O. J. Z. (2008). A multicenter, randomized clinical trial of a cognitive remediation program for childhood survivors of a pediatric malignancy. *Journal of Consulting and Clinical Psychology, 76*, 367–378.

Chan, D. Y. K., & Fong, K. N. K. (2011). The effects of problem-solving skills training based on metacognitive principles for children with acquired brain injury attending mainstream schools: A controlled clinical trial. *Disability and Rehabilitation, 33*, 2023–2032.

Cohen, J. (1988). *Statistical analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Conners, C. K. (1997). *Conners’ Rating Scales-Revised*. North Tonawanda, NY: Multi-Health Systems.
DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*, 177–188.

Gibson, B. S., Gondoli, D. M., Johnson, A. C., Steeger, C. M., Dobrzenski, B. A., & Morrissey, R. A. (2011). Component analysis of verbal versus spatial working memory training in adolescents with ADHD: A randomized, controlled trial. *Child Neuropsychology, 17*, 546–563.

Gray, S. A., Chaban, P., Martinussen, R., Goldberg, R., Gotlieb, H., Kronitz, R., ... Tannock, R. (2012). Effects of a computerized working memory training program on working memory, attention, and academics in adolescents with severe LD and comorbid ADHD: A randomized controlled trial. *Journal of Child Psychology and Psychiatry, 53*, 1277–1284.

Green, C. T., Long, D. L., Green, D., Iosif, A., Dixon, J. F., Miller, M. R., ... Schweitzer, J. B. (2012). Will working memory training generalize to improve off-task behavior in children with attention-deficit/hyperactivity disorder? *Neurotherapeutics, 9*, 639–648.

Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., ... Schunemann, H. J. (2011). GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology, 64*, 383–394.

Higgins, J. P. T., & Green, S. (Eds.). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*. The Cochrane Collaboration, 2011. Retrieved from www.cochrane-handbook.org

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*, 1539–1558.

Johnstone, S. J., Roodenrys, S., Blackman, R., Johnston, E., Loveday, K., Mantz, S., & Barratt, M. F. (2012). Neurocognitive training for children with and without AD/HD. *Attention Deficit Hyperactivity Disorder, 4*, 11–23.

Johnstone, S. J., Roodenrys, S., Phillips, E., Watt, A. J., & Mantz, S. (2010). A pilot study of combined working memory and inhibition training for children with AD/HD. *Attention Deficit Hyperactivity Disorder, 2*, 31–42.

Kennyard, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*, 983–997.

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlstrom, K., ... Westerberg, H. (2005). Computerized training of working memory in children with ADHD: A randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry, 44*, 177–186.

Laatto, L., Harrington, D., Hotz, G., Marcatuono, J., Mozzoni, M. P., Walsh, V., & Hersey, K. P. (2007). An evidence-based review of cognitive and behavioral rehabilitation treatment studies in children with acquired brain injury. *Journal of Head Trauma Rehabilitation, 22*, 248–256.

McCauley, S. R., Wilde, E. A., Anderson, V. A., Bedell, G., Beers, S. R., Campbell, T. F., ... Pediatric TBI Outcomes Workgroup. (2012). Recommendations for the use of common outcome measures in pediatric traumatic brain injury research. *Journal of Neurotrauma, 29*, 678–705.

McCormick, A., Aubut, J., Gnanakumar, V., Curiale, A., & Marshall, S. (2012). Pediatric interventions in acquired brain injury rehabilitation, ERABI. London, Ontario.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective: A meta-analytic review. *Developmental Psychology, 49*, 270–291.

Ozonoff, S. (2010). Autism spectrum disorders. In K.O. Yeates, M. D. Ris, H. G. Taylor, & B. F. Pennington (Eds.), *Pediatric neuropsychology: Research, theory, and practice* (2nd ed., pp. 418–446). New York, NY: Guilford Press.

Rapport, M. D., Orban, S. A., Koller, M. J., & Friedman, L. M. (2013). Do programs designed to train working memory, other executive functions, and attention benefit children with ADHD? A meta-analytic review of cognitive, academic, and behavioral outcomes. *Clinical Psychology Review, 33*, 1237–1252.

Ris, M. D., & Abbey, R. (2010). Pediatric brain tumors. In K. O. Yeates, M. D. Ris, H. G. Taylor, & B. F. Pennington (Eds.), *Pediatric neuropsychology: Research, theory, and practice* (2nd ed., pp. 92–111). New York: Guilford Press.

Ross, K. A., Dorris, L., & McMillan, T. (2011). A systematic review of psychological interventions to alleviate cognitive and psychosocial problems in children with acquired brain injury. *Developmental Medicine and Child Neurology, 53*, 692–701.

Schulz, K. F., Altman, D. G., & Moher, D.; CONSORT Group (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *PLoS Medicine, 7*, e1000251. doi:10.1371/journal.pmed.1000251

Shalev, L., Tsal, Y., & Mevorach, C. (2007). Computerized Progressive Attentional Training (CPAT) program: Effective direct intervention for
children with ADHD. *Child Neuropsychology*, 13, 382–388.

Slomine, B., & Locascio, G. (2009). Cognitive rehabilitation for children with acquired brain injury. *Developmental Disabilities Research Reviews*, 15, 133–143.

Solan, H. A., Shelley-Tremblay, J., Ficarra, A., Silverman, M., & Larson, S. (2003). Effect of attention therapy on reading comprehension. *Journal of Learning Disabilities*, 36, 556–563.

Toplak, M. E., Connors, L., Shuster, J., Knezevic, B., & Parks, S. (2008). Review of cognitive, cognitive-behavioral, and neural-based interventions for Attention-Deficit/Hyperactivity Disorder (ADHD). *Clinical Psychology Review*, 28, 801–823.

Van’t Hooft, I., Andersson, K., Bergman, B., Sejersen, T., von Wendt, L., & Bartfai, A. (2007). Sustained favorable effects of cognitive training in children with acquired brain injuries. *NeuroRehabilitation*, 22, 109–116.

Westerveld, M. (2010). Childhood epilepsy. In K. O. Yeates, M. D. Ris, H. G. Taylor, & B. F. Pennington (Eds.), *Pediatric neuropsychology: Research, theory, and practice* (2nd ed., pp. 71–91). New York: Guilford Press.

Willcutt, E. G. (2010). Pediatric brain tumors. In K. O. Yeates, M. D. Ris, H. G. Taylor, & B. F. Pennington (Eds.), *Pediatric neuropsychology: Research, theory, and practice* (2nd ed., pp. 393–417). New York: Guilford Press.

Wolfe, K. R., Madan-Swain, A., & Kana, R. K. (2012). Executive dysfunction in pediatric posterior fossa tumor survivors: A systematic literature review of neurocognitive deficits and interventions. *Developmental Neuropsychology*, 37, 153–175.

Yeates, K. O. (2010). Traumatic brain injury. In K. O. Yeates, M. D. Ris, H. G. Taylor, & B. F. Pennington (Eds.), *Pediatric neuropsychology: Research, theory, and practice* (2nd ed., pp. 112–146). New York, NY: Guilford Press.