Neurocognition after paediatric heart surgery: a systematic review and meta-analysis

Caroline Sterken,1 Jurgen Lemiere,2,3 Ilse Vanhorebeek,1 Greet Van den Berghe,1 Dieter Mesotten1

ABSTRACT

Children with congenital heart disease (CHD) often experience difficulties in academic and daily functioning, which have been associated with intelligence and neurocognitive skills, including executive functions (EFs), attention and memory. We report the neurocognitive data of children with CHD who were included in the Leuven glucose control trial (LGC trial). Through a systematic review and meta-analysis, we aimed to find which neurocognitive functions are most consistently and prominently affected. 365 children with CHD and 216 healthy control children underwent extensive neurocognitive testing in the LGC trial. A comprehensive search of electronic databases PubMed, EMBASE and Cochrane was conducted for studies measuring intelligence, EFs, attention and memory in children who underwent heart surgery for CHD. Standardised mean differences (SMDs) between the CHD group and a healthy control group were calculated for these neurocognitive functions. LGC trial data were included in the meta-analysis. Twelve studies with a healthy control group were included in the meta-analysis, involving 647 patients with CHD and 633 controls. The CHD group (median age 7.35 years at testing) had worse scores than healthy control children, for all investigated neurocognitive functions. A medium SMD was found for intelligence (SMD=−0.53 (95% CI −0.68 to −0.38), p<0.00001). Alertness, an attentional function, was also consistently poorer in the CHD group. Memory was less affected, while EF had a medium SMD with large heterogeneity. Children with CHD risk displayed lower performance on intelligence and alertness assessment, which may contribute to difficulties in daily life and school. Heterogeneity in neurocognitive assessment and small sizes in most studies limit the interpretation.

Trial registration number: clinicaltrials.gov Identifier NCT00214916.

INTRODUCTION

As the survival rate of children with congenital heart disease (CHD) continues to improve thanks to medical advances,1 public interest and research have been focusing more on how children with CHD survive.2 Children with CHD may experience difficulties in daily life3 and academic functioning,4 which may persist into adulthood.5 In this systematic review and meta-analysis, we aim to examine intelligence and specific neurocognitive skills, especially executive functions (EFs), as important predictors for academic and daily functioning in children with CHD after heart surgery.

Intelligence scores (IQs) are good predictors of academic performance.6 However, children with CHD often have intelligence scores within average range compared with population norms.7 Research in other paediatric conditions, such as traumatic brain injury, has demonstrated that intelligence scores are rough measures and fall short in detecting more specific neurocognitive skills, such as EFs.8 EFs cover a variety of cognitive functions, such as planning, organisation, flexibility, cognitive control and working memory. These are essential in many domains of daily life9 and contribute to academic performance.10 11 Other neurocognitive skills, such as memory and attention, which are inter-related with EFs, also contribute to academic performance.9

There is growing evidence that brain development in children with CHD can differ from normal brain development. This misdevelopment may even start prenatally due to impaired cerebral blood flow.12 Postnatally, infants and older children with CHD may have preoperative13 and postoperative white matter abnormalities,14 which may relate to worse neurocognitive outcome in children with CHD.15 Surgery seems to impact neurocognitive outcome in CHD as well.15 Considering the vulnerability of their brain, neurocognitive functions may be worse in children with CHD than those in healthy control children without CHD. Some relevant reviews have discussed the importance of these neurocognitive functions.2 16 17

1Division of Cellular and Molecular Medicine, Clinical Department and Laboratory of Intensive Care Medicine, KU Leuven, Leuven, Belgium
2Child and Adolescent Psychiatry, University Hospitals Leuven, Leuven, Belgium
3Paediatric Haematology-Oncology, University Hospitals Leuven, Leuven, Belgium

Correspondence to
Dr Dieter Mesotten, dieter.mesotten@med.kuleuven.be

Received 18 February 2015
Revised 18 September 2015
Accepted 29 September 2015
However, to our knowledge, the outcome data of specific neurocognitive functions have not been analysed systematically in children with CHD. One systematic review and one meta-analysis examined intellectual outcome but not more specific neurocognitive skills in children with CHD after heart surgery.

A large randomised controlled trial (n=700), in which neurocognitive development of children was assessed 4 years after critical illness and treatment with tight glucose control, has recently been completed. The results demonstrated that tight glucose control in critically ill children improved motor coordination and cognitive flexibility in comparison with children in whom blood glucose levels up to 215 mg/dL were tolerated.

Seventy-five per cent of the study population in this Leuven glucose control trial (LGC trial) underwent heart surgery for congenital heart defects. Neurocognitive data of the heart surgery subgroup have not yet been analysed.

Thus, the first aim of this paper was to report the neurocognitive data of this large cohort of children with heart surgery for CHD, included in the LGC trial, and healthy controls. The second aim was to carry out a systematic review and meta-analysis for intelligence, EF, attention and memory in children with CHD after heart surgery. We hypothesised that specific neurocognitive skills such as EFs are more impaired than is intelligence, in children with CHD.

METHODS

Analysis of the LGC trial data

Data of all children who underwent neurocognitive testing and for whom a full-scale IQ was available, were analysed (children with CHD n=361, healthy controls n=215). For four children of the CHD group and one child of the control group, a full-scale IQ could not be calculated. Baseline neurocognitive data, at the time patients were included into the LGC trial, were not available. Demographic, clinical and neurocognitive data are reported as numbers and percentages, or as median and IQR. Because of imbalance between the CHD and control group for gender, presence of a syndrome, socioeconomic status and age at follow-up, propensity score matching was performed using IBM SPSS Statistics V.22.0.0.1 and R statistical software V.2.15.3. For more details on propensity score matching, we refer to eMethods 1. Demographic, clinical and neurocognitive data of the tested and matched population were further analysed using $\chi^2$ test for dichotomous variables and unpaired non-parametric Wilcoxon rank-sum tests for continuous variables (JMP V.11.2.0 (SAS Institute)). The details of the study protocol and neurocognitive test battery have previously been reported.

Systematic review and meta-analysis

Data sources and searches

A comprehensive search of electronic databases PubMed, EMBASE and Cochrane was conducted for studies published between the beginning of each database and December 2014. Each search strategy consisted of four major parts: cognition, heart, child and tests. Both index language terms (MeSH, Emtree) and keywords were used in every part of the search strategy (see eMethods 2). We also manually screened reference lists of studies identified through database search.

Study selection

Two selection criteria were premised for title and abstract screening. First, the study population consisted of infants, children, adolescents and/or young adults (<24 years old) with CHD. Second, they needed to have, at least, intelligence testing with an indication of a specific measure of EF, memory or attention, or a broader neurocognitive assessment. For the full-text screening, we introduced two more selection criteria. First, the participants had neurocognitive testing after heart surgery or an interventional cardiac procedure. Studies that tested participants in the first six postoperative months were excluded to avoid interference with the acute medical phase. Second, at least one test that measured EF, memory and/or attention was needed.

Data extraction

First, we gathered data on the sample size, age at surgery and age at testing of the tested groups. We also investigated whether the same study population was tested in already included studies of the same research group. Furthermore, for comparison of the CHD group versus a healthy control group, we collected data necessary to quantify differences between the CHD and healthy control group. The collected data are summarised in box 1. In case of missing data for the quantitative analysis, authors were contacted.

Data analysis

As appropriate, mean (SD) or median (IQR) sample size, age at surgery and age at testing were calculated for all studies. Risk of bias of all included studies was assessed by two raters (CS and JL) independently by means of a modified version of the Downs and Black...
checklist.32 Six items that focus on RCTs (items 4, 8, 14, 19, 23, 24) were omitted, leading to a maximum total score of 22. In case of different scores by the two raters, a consensus was reached through discussion. Inter-rater reliability was good (Spearman’s r 0.722; p=0.008). Depending on the number of included studies, mean (SD) or median (IQR) total risk of bias and subscores were calculated for all studies.

Analyses were performed on neurocognitive data of studies that used a healthy control group. To combine these continuous data measured by different instruments, effect sizes, that is, standardised mean differences (SMDs), were calculated in Review Manager V.5.2 by means of Inverse Variance, random effects analysis (http://ims.cochrane.org/revman). SMDs were calculated for the following neurocognitive functions: intelligence, alertness (attentional function), memory and inhibition (EF). Neurocognitive data of the LGC trial were included in the SMD calculation. A two-sided p value <0.05 was considered statistically significant. SMDs were classified according to Cohen’s guidelines: d=0.2 defined as small, d=0.5 as medium and d=0.8 as large.33 For clinical interpretation, the overall SMD for intelligence scores was multiplied by the typical SD of the normal IQ distribution (mean 100±SD 15). The F statistic was used to evaluate heterogeneity. F² ≥50% was defined as substantial heterogeneity. Funnel plots were also made in Review Manager when at least 10 studies were available, to explore small-study effects. For more details on study selection and data analysis, we refer to eMethods 1.

RESULTS
Analysis of the LGC trial data
Demographic and clinical data of tested and matched postheart surgery population and healthy controls are presented in table 1. Neurocognitive data are presented in table 2. Children of the tested postheart surgery population have worse scores for intelligence, visual-motor integration and all measures of alertness, motor coordination, inhibition, flexibility (except for ΔN of errors), memory and behaviour, compared with healthy controls. After propensity score matching, the CHD group has worse scores for intelligence, visual-motor integration, motor coordination (alternating taps), inhibition (ΔN of errors), memory (verbal working memory and immediate memory) and behaviour (internalising and total problems).

Systematic review and meta-analysis
Flow diagram
Figure 1 shows the article screening phases and the reasons for exclusion, according to the PRISMA guidelines (Preferred Reporting Items for Systematic reviews and Meta-Analysis).34 An overview of the 12 included studies is presented in eTable 1.

Sample size, and age at surgery and at testing
The median sample size of the CHD groups in all 12 studies included in the meta-analysis was 31 (IQR 18–43). The median sample size of the healthy control groups was 33 (IQR 20–42). The median age at surgery was 1.25 years (IQR 0.7–3.85) and the median age at testing was 7.35 years (IQR 5.70–8.37).

Quality assessment
The median (IQR) risk of bias of all 12 included studies was 16 (IQR 15–17). Median score for reporting was 8 (IQR 7–8) with a maximum score of 9. Median score for external validity was 2 (IQR 1–2) with a maximum score of 3 and median score for internal validity was 7 (IQR 5–7) with a maximum score of 9. Only 1 out of 12 studies mentioned that the study had sufficient power to detect a clinically relevant difference.

Meta-analyses: CHD group versus healthy controls
Total SMDs for all neurocognitive domains indicate lower scores for the CHD group. A medium SMD with low heterogeneity was found for intelligence (SMD=−0.53 (95% CI −0.68 to −0.38), p<0.00001, F²=32%; figure 2), indicating a drop of 0.53 times the SD of the normal IQ distribution: −8 IQ points. A medium SMD was also found for alertness non-reaction time (SMD=−0.47 (95% CI −0.67 to −0.27), p<0.00001, F²=19%; figure 3) and a smaller SMD for alertness reaction time (SMD=−0.25 (95% CI 0.08 to 0.42), p=0.004, F²=0%; efigure 1). Verbal memory showed a smaller SMD (SMD=−0.35 (95% CI −0.54 to −0.15), p<0.00001, F²=0%; figure 4), while non-verbal memory did not differ between children with CHD and healthy controls (efigure 2). EF reaction time, examining the inhibition function, had the largest SMD, but a high level of heterogeneity (SMD=0.57 (95% CI 0.10 to 1.04), p=0.02, F²=80%; figure 5). EF non-reaction time for inhibition also showed a medium SMD but with low heterogeneity (SMD=−0.51 (95% CI −0.74 to −0.29), p<0.00001, F²=0%; efigure 3). The funnel plot for intelligence was slightly asymmetrical (efigure 4). Owing to the limited availability of studies on the other neurocognitive functions, funnel plots could not be built for alertness, memory and inhibition.

DISCUSSION
This meta-analysis and the results of the LGC trial provide sound evidence that not only intelligence but also more specific neurocognitive functions are impaired in children with CHD who underwent heart surgery, compared with healthy controls. Contrary to our hypothesis, intelligence, EFs and alertness seem to be equally affected.

Intelligence
The SMD of −0.53 (95% CI −0.68 to −0.38) in the meta-analysis and the effect size of 8 IQ points in both
| Table 1  Demographic and clinical data of tested and propensity score-matched postheart surgery population and healthy controls in the LGC trial |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| **Tested population** | **Propensity score-matched population** | **Healthy controls** | **Healthy controls** |
| Tested—post heart surgery population (N=361) | Tested—post heart surgery population (N=167) | Healthy controls (N=215) | Healthy controls (N=167) |
| Caucasian race* | 343 (95.01) | 330 (91.41) | 211 (95.01) | 205 (95.01) |
| Exclusive European* | 296 (81.99) | 201 (93.49) | 166 (80.39) | 163 (93.20) |
| Male sex* | 205 (56.79) | 144 (86.23) | 93 (43.26) | 81 (48.50) |
| Age at randomisation, years† | 0.76 (0.22–4.1) | 2.11 (0.34–4.77) | NA | NA |
| Type of congenital heart disease* | Obstructive pathology | 79 (21.88) | 37 (22.96) | NA | NA |
| Left-right shunt | 121 (33.51) | 122 (33.79) | NA | NA |
| Cyanotic and not univentricular | 70 (19.39) | 59 (35.32) | NA | NA |
| Cyanotic and univentricular | 122 (33.79) | 12 (7.44) | NA | NA |
| Other | 11 (3.04) | 5 (2.99) | NA | NA |
| RACHS classification* | 1 or 2 | 168 (47.72) | 79 (48.17) | NA | NA |
| 3 or 4 | 175 (49.71) | 84 (51.21) | NA | NA |
| 5 or 6 | 9 (2.55) | 1 (0.60) | NA | NA |
| Syndrome, at randomisation* | PELOD first 24 h in ICU† | 69 (19.11) | 16 (7.44) | 11 (2–12) | 16 (9.58) |
| Socioeconomic status score† | 42.5 (29–54) | 39.5 (29–50) | 35 (24–48.5) | 39.5 (29–52.5) |
| At follow-up† | Height, cm | 107 (103–126) | 122 (108–151) | 116 (104–131) | 117 (107–139) |
| Weight, kg | 18 (15–24) | 22 (18–40) | 20 (16–29) | 21 (18–34) |
| Head circumference, cm | 50.7 (49.2–52.5) | 52 (50.8–54) | 51 (49.5–53) | 51.8 (50.5–53.4) |
| Age, year | 4.67 (4.14–7.93) | 6.75 (4.14–7.93) | 6.02 (4.21–8.78) | 5.91 (4.58–9.07) |

*Numbers and percentages.
†Median (IQR).
ICU, intensive care unit; LGC, Leuven glucose control; NA, not applicable; PELOD, paediatric logistic organ dysfunction; RACHS, risk adjustment for congenital heart surgery.42

### Table 2  Results of neurocognitive test battery in tested and propensity score-matched postheart surgery population and healthy controls in the LGC trial

<table>
<thead>
<tr>
<th></th>
<th>Tested population</th>
<th></th>
<th>Propensity score-matched population</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Post heart surgery</td>
<td>Healthy controls</td>
<td>p Value</td>
<td>Post heart surgery</td>
</tr>
<tr>
<td></td>
<td>population (N=361)</td>
<td>(N=215)</td>
<td></td>
<td>population (N=167)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical neurological evaluation score (range 0–8)*</td>
<td>1 (0–2)</td>
<td>0 (0–1)</td>
<td>&lt;0.001</td>
<td>1 (0–2)</td>
</tr>
<tr>
<td>Full-scale IQ</td>
<td>90 (75–100)</td>
<td>103 (91–111)</td>
<td>&lt;0.001</td>
<td>92 (78–103)</td>
</tr>
<tr>
<td>Verbal IQ</td>
<td>90.5 (76.2–101)</td>
<td>102 (92–114)</td>
<td>&lt;0.001</td>
<td>94 (79.5–104)</td>
</tr>
<tr>
<td>Performance IQ</td>
<td>89 (77–101)</td>
<td>103 (92–112)</td>
<td>&lt;0.001</td>
<td>91 (77–102)</td>
</tr>
<tr>
<td>Visual-motor integration (range 0.9–20)†</td>
<td>9 (7–10)</td>
<td>10 (8–12)</td>
<td>&lt;0.001</td>
<td>9 (7–10)</td>
</tr>
<tr>
<td>Intelligence (range of possible scores, 45–155)†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction time dominant hand, ms</td>
<td>691 (447–982)</td>
<td>481 (320–700)</td>
<td>&lt;0.001</td>
<td>566 (374–829)</td>
</tr>
<tr>
<td>Within-patient SD of repeated tests</td>
<td>404 (143–642)</td>
<td>165 (83–383)</td>
<td>&lt;0.001</td>
<td>228 (115–535)</td>
</tr>
<tr>
<td>Reaction time non-dominant hand, ms</td>
<td>697 (436–968)</td>
<td>499 (326–721)</td>
<td>&lt;0.001</td>
<td>543 (374–791)</td>
</tr>
<tr>
<td>Within-patient SD of repeated tests</td>
<td>336 (158–611)</td>
<td>192 (87–379)</td>
<td>&lt;0.001</td>
<td>221 (110–482)</td>
</tr>
<tr>
<td>Motor coordination (Number of taps in 10 s)†</td>
<td>28 (22–38)</td>
<td>35 (25–46)</td>
<td>&lt;0.001</td>
<td>31 (22–42)</td>
</tr>
<tr>
<td>Number of unimanual taps dominant hand</td>
<td>23 (17–33)</td>
<td>29 (21–43)</td>
<td>&lt;0.001</td>
<td>25 (18–35)</td>
</tr>
<tr>
<td>Number of unimanual taps non-dominant hand</td>
<td>8 (2–18)</td>
<td>13 (5–30)</td>
<td>&lt;0.001</td>
<td>8 (2–20)</td>
</tr>
<tr>
<td>Number of valid alternating taps</td>
<td>16 (8–24)</td>
<td>21 (12–31)</td>
<td>&lt;0.001</td>
<td>18 (11–27)</td>
</tr>
<tr>
<td>Number of valid synchronous taps</td>
<td>2 (0–4)</td>
<td>4 (2–8)</td>
<td>&lt;0.001</td>
<td>1 (0–3)</td>
</tr>
<tr>
<td>Inhibition and flexibility*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔReaction time (inhibition), ms</td>
<td>313 (120–536)</td>
<td>200 (79–485)</td>
<td>0.008</td>
<td>261 (98.7–438)</td>
</tr>
<tr>
<td>ΔNumber of errors (inhibition)</td>
<td>2 (0–3)</td>
<td>1 (0–2)</td>
<td>0.002</td>
<td>1.5 (0–3)</td>
</tr>
<tr>
<td>ΔReaction time (flexibility), ms</td>
<td>637 (367–878)</td>
<td>550 (283–798)</td>
<td>0.043</td>
<td>603 (330–848)</td>
</tr>
<tr>
<td>ΔNumber of errors (flexibility)</td>
<td>2 (0–4)</td>
<td>4 (2–8)</td>
<td>0.243</td>
<td>2 (0–4)</td>
</tr>
<tr>
<td>Memory†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Numbers (range 1–19): memory span (forward)</td>
<td>8 (5–9)</td>
<td>9 (7–11)</td>
<td>&lt;0.001</td>
<td>8 (6–11)</td>
</tr>
<tr>
<td>Numbers (range 1–19): working memory (backward)</td>
<td>9 (6–11)</td>
<td>10 (8–13)</td>
<td>&lt;0.001</td>
<td>9.5 (6.2–12)</td>
</tr>
<tr>
<td>Word pairs (proportion of correct responses): learning</td>
<td>0.45 (0.33–0.53)</td>
<td>0.5 (0.38–0.66)</td>
<td>0.001</td>
<td>0.46 (0.36–0.57)</td>
</tr>
<tr>
<td>Word pairs (proportion of correct responses): immediate memory</td>
<td>0.4 (0.2–0.5)</td>
<td>0.5 (0.35–0.64)</td>
<td>&lt;0.001</td>
<td>0.4 (0.23–0.5)</td>
</tr>
<tr>
<td>Word pairs (proportion of correct responses): delayed memory</td>
<td>0.3 (0.2–0.4)</td>
<td>0.4 (0.3–0.5)</td>
<td>&lt;0.001</td>
<td>0.35 (0.21–0.42)</td>
</tr>
<tr>
<td>Word pairs (proportion of correct responses): recognition</td>
<td>0.96 (0.9–1)</td>
<td>1 (0.95–1)</td>
<td>0.011</td>
<td>0.97 (0.9–1)</td>
</tr>
<tr>
<td>Non-verbal, visual-spatial (proportion of correct responses)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pictures: memory span</td>
<td>0.83 (0.71–0.89)</td>
<td>0.89 (0.8–0.93)</td>
<td>&lt;0.001</td>
<td>0.84 (0.73–0.9)</td>
</tr>
<tr>
<td>Dots: learning</td>
<td>0.83 (0.66–0.88)</td>
<td>0.88 (0.83–0.94)</td>
<td>&lt;0.001</td>
<td>0.87 (0.70–0.94)</td>
</tr>
<tr>
<td>Dots: immediate memory</td>
<td>0.83 (0.62–0.81)</td>
<td>1 (0.75–1)</td>
<td>&lt;0.001</td>
<td>0.87 (0.66–1)</td>
</tr>
<tr>
<td>Dots: delayed memory</td>
<td>0.83 (0.56–0.81)</td>
<td>1 (0.75–1)</td>
<td>&lt;0.001</td>
<td>0.83 (0.66–1)</td>
</tr>
<tr>
<td>Learning index (range 50–150)</td>
<td>93 (84–103)</td>
<td>101 (90–109)</td>
<td>&lt;0.001</td>
<td>96 (87–103)</td>
</tr>
</tbody>
</table>

Continued
the meta-analysis and LGC trial are of similar magnitude to those in prematurely born children,\textsuperscript{35} and are in line with the meta-analysis of Karsdorp \textit{et al.}\textsuperscript{19} Despite the worse intelligence scores for the CHD group compared with healthy controls, the mean IQ scores of the children with CHD were still within the ‘normal’ average range, when compared with population norms. Nevertheless, these IQ differences can have consequences in a learning environment.\textsuperscript{36} Population norms might insufficiently capture the dynamic nature of development. A well-matched healthy control group of typically developing children without CHD may offer a more representative reflection of normal variation between and within typically developing children\textsuperscript{37,38} who grow up in the same time period as children with CHD.

\textbf{Alertness, inhibition and memory}

Similarly to intelligence, alertness was also consistently impaired across the available studies in the meta-analysis. The alerting system is one of the attentional networks in addition to the orienting and executive attention network.\textsuperscript{39} It is responsible for achieving and maintaining a state of high sensitivity to incoming information. An efficient alerting system is pivotal for other more complex cognitive functions, such as EF. Additionally, detecting problems of alertness has an important clinical consequence because improving alertness is considered essential for cognitive rehabilitation.

Recently, more consideration has been given to impairments in EF in children with CHD.\textsuperscript{21,61,7} The LGC trial exactly showed that, in critically ill children, the more complex EF, flexibility, was improved by tight glucose control, almost to the levels of healthy controls\textsuperscript{20} However, from this meta-analysis, we have to conclude that studies examining these specific neurocognitive skills are scarce. The few studies included in this meta-analysis involved several specific subgroups of CHD with only a limited number of patients in each subgroup. This resulted in low statistical power and precision. Also, the use of different tests for EF, adding to the ascertained heterogeneity, may have had an impact on low precision.

Nevertheless, the SMD (0.57) for EF reaction time was in line with the deficit in intelligence (0.53). It indicates that children with CHD react slower when performing specific EF tasks measuring inhibition. Children with CHD from the LGC trial showed the smallest SMD in EF reaction time.\textsuperscript{20} This may be attributed to the larger size of the study and to the fact that the EF reaction time was corrected for baseline speed and the propensity score matching. The CHD group of the LGC trial also made more errors during an inhibition test. This meta-analysis confirmed the findings from previous reviews,\textsuperscript{21,16,17} contending that the risk of memory deficits in children with CHD may be lower. However, results from the LGC trial showed worse performance for immediate memory and working memory of verbal information in the CHD group. The use of pooled memory scores in the

\begin{table}
\centering
\begin{tabular}{|c|c|c|c|}
\hline
\textbf{Tested population} & \textbf{Propensity score-matched population} & \textbf{Propensity score-matched population} & \textbf{Propensity score-matched population} \\
\hline
\textbf{Tested—post heart surgery population (N=361)} & \textbf{Healthy controls (N=215)} & \textbf{p Value} & \textbf{p Value} \\
\hline
\textbf{Behaviour (by proxy), T-score 1} & \textbf{CBCL-internalising problems (range 29–100)} & \textbf{52 (45–61)} & \textbf{48 (41–58)} & \textbf{<0.001} & \textbf{<0.001} \\
\hline
\textbf{CBCL-externalising problems (range 28–100)} & \textbf{52 (43–56)} & \textbf{51 (43–56)} & \textbf{0.046} & \textbf{<0.001} \\
\hline
\textbf{CBCL-total problems (range 24–100)} & \textbf{52 (45–59)} & \textbf{51 (43–56)} & \textbf{<0.001} & \textbf{<0.001} \\
\hline
\hline
\textbf{Tested—post heart surgery population (N=167)} & \textbf{Healthy controls (N=167)} & \textbf{p Value} & \textbf{p Value} \\
\hline
\textbf{CBCL-internalising problems (range 29–100)} & \textbf{49 (41–58)} & \textbf{51 (43–56)} & \textbf{<0.001} & \textbf{<0.001} \\
\hline
\textbf{CBCL-externalising problems (range 28–100)} & \textbf{47 (40–59)} & \textbf{49 (41–59)} & \textbf{0.221} & \textbf{0.006} \\
\hline
\textbf{CBCL-total problems (range 24–100)} & \textbf{48 (40–59)} & \textbf{47 (40–55)} & \textbf{0.049} & \textbf{0.066} \\
\hline
\hline
\textbf{Median (IQR) are reported. *Higher scores reflect worse performance. †Higher scores reflect better performance. CBCL, Child Behavior Checklist.}
\end{tabular}
\end{table}
A meta-analysis may have hidden specific memory deficits. Children may also have been too young to detect differences in tests, which examine functions that are continuously developing through childhood.

The aforementioned indicates that the difficulties that children with CHD experience when performing IQ tests and complex tasks requiring EFs may be explained by a basic alertness deficit. This has also been found in other paediatric populations and may be linked with white matter changes. Alternatively, other EFs, such as flexibility or working memory, may preferentially be affected.

**Figure 1** Flow diagram of Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA). 1Search strategy in eMethods 1. 2Manuscripts could be excluded for more than one reason (CHD, congenital heart disease).
Strengths
This systematic review and meta-analysis has several strengths. Exploring the effects of CHD on not only intelligence but also on EFs, attention and memory in the same patients has been a new avenue in the assessment of neurocognitive function in children with CHD. This allowed us to compare the impact of CHD on intelligence with the impact of CHD on specific neurocognitive skills. The studies that included a healthy control group could be combined for the analysis of a SMD, despite the use of different tests. Therefore, the findings are fairly robust. The analysis of the LGC trial data offers a valuable contribution of neurocognitive data from a large CHD and healthy control group, thereby increasing power and precision in the meta-analysis. Certainly, propensity score matching for relevant factors in neurocognitive development improved the stringency and reliability of the analyses.

Limitations
The meta-analysis has a few inherent limitations though. First, attention and EFs involve several functions, which may interact. Although the lower aggregated score on intelligence is clinically relevant, it is not clear whether the poorer results for attention and EF have any clinical impact, because universal definitions and test protocols are lacking. Unfortunately, the number of studies assessing different aspects of attention and EF was low. As a result, only one attentional function and one EF could be examined in this meta-analysis. Therefore, no conclusions can be drawn for other attention components and EFs. Future studies thus ought to use comparable test batteries for neurocognitive function assessments. Second, the very small sample sizes of the CHD and control groups, partially due to the separate reporting of a high variation in CHD diagnoses, reduced statistical power and precision both at the individual study level and at meta-analysis level. Because of the separate reporting of CHD subgroups, neurocognitive data of healthy controls were sometimes included twice in the forest plots. The slightly asymmetrical funnel plot of intelligence reflects a possible publication bias and may have overestimated the effect size. Future research should pay more attention to the statistical powering of

Figure 2  Forest plot of intelligence (author/year/journal—CHD type). TGA, d-transposition of the great arteries; VSD, ventricular septal defect; TOF, tetralogy of Fallot; ASD, atrial septal defect; CHD, congenital heart disease; ECMO, extracorporeal membrane oxygenation; V.d.Rijken, Van der Rijken; cath, catheterisation; surg, surgery; CA, circulatory arrest; CPB, cardiopulmonary bypass.

Figure 3  Forest plot of attention (Att; alertness) non-reaction time (author/year/journal—CHD-type/attention measure). Att1, mean accuracy (%) Attention Network Test; Att2, visual attention (NEuroPSYchological Assessment, NEPSY); Att3, auditory attention (NEPSY); Att4, processing speed (Wechsler Intelligence Scale for Children, WISC-IV); Att5, delay task, vigilance hits; Att6, speed of information (British Ability Scales, BAS); VSD, ventricular septal defect; TOF, tetralogy of Fallot; ASD, atrial septal defect; CHD, congenital heart disease; ECMO, extracorporeal membrane oxygenation.
effects on neurocognitive outcome. Third, the exclusion of six non-English studies (one Italian, one Chinese, one French, one Spanish and two German) might have raised a language bias. However, research has shown no evidence of systematic bias from the use of language restrictions in systematic review-based meta-analyses.41

Fourth, this meta-analysis cannot draw any conclusion on the interaction between effect of IQ and alertness, EF and memory. At least to exclude the fact that effects on EFs and memory are not entirely explained by a gen-

eralised or more basic impairment, one ought to examine IQ and alertness in addition to the other functions. Finally, studies on EFs, attention and memory outcomes in adolescence are lacking, indicating the need for long-term longitudinal follow-up studies to the level of secondary education,18 when academic difficulties can appear. Adolescence is particularly essential for the late maturation of the prefrontal cortex and EFs.37

Understanding the impact of these prefrontal cortex functions on daily life and school functioning is clinically relevant, because a better understanding and early detection of deficits may improve the daily functioning of children with CHD.2

CONCLUSIONS

Children with CHD who have undergone heart surgery have consistently worse performance of intelligence and alertness. In this meta-analysis, memory appears to be less affected. The effect of CHD on EF in young children cannot be reliably assessed due to poor standardisation of the testing methodology. Larger, more standardised, longitudinal long-term follow-up studies of specific neurocognitive skills in a large group of children with CHD and a matched healthy control group are necessary for a better understanding of neurocognitive deficits, and their impact on daily life and school functioning.

Acknowledgements The authors thank Jens De Groot (Biomedical Library, Catholic University Leuven) for his help with the search strategies.

Contributors CS and JL conceptualised and designed the study, carried out analyses, drafted the initial manuscript and approved the final manuscript as submitted. IV carried out analyses, and reviewed and approved the final manuscript as submitted. GVdB conceptualised and designed the study, carried out analyses, and reviewed and approved the final manuscript as submitted. DM conceptualised and designed the study, reviewed and revised the manuscript, and approved the final manuscript as submitted.

Funding This study was funded by the Flemish Agency for Innovation by Science and Technology (IWT/070695/TBM). DM holds a FWO-Flanders Senior Clinical Investigator Fellowship. GVdB via the University of Leuven receives structural research financing via the Methusalem programme, funded by the Flemish government, and holds an ERC advanced grant from the EU-FP7 (2012-321670).

Competing interests None declared.

Patient consent Obtained.

Ethics approval Institutional Review Board.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license,


Figure 4 Forest plot of verbal memory (Mem; author/year/journal—CHD-type/memory measure). Mem1, sum of immediate and delayed verbal memory (Children’s Memory Scale, Word Pairs, proportion correct responses); Mem2, narrative memory (NEuroPSYchological Assessment, NEPSY); Mem3, verbal memory (Wide Range Assessment of Memory and Learning, WRAML); ASD, atrial septal defect; CHD, congenital heart disease; ECMO, extracorporeal membrane oxygenation.

Figure 5 Forest plot of EF (inhibition) RT (author/year/journal—CHD-type/EF measure). EF, executive function; RT, reaction time; EF-test1, incongruent stroop; EF-test2, stroop; EF-test3, conflict (Attention Network Test); EF-test4, inhibition (Amsterdam Neuropsychological Tasks); TGA, d-transposition of the great arteries; VSD, ventricular septal defect; TOF, tetralogy of Fallot; CHD, congenital heart disease.
REFERENCES


