

*Cardiology in the Young* (2018), 28, 571–581 © Cambridge University Press, 2018. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.  
doi:10.1017/S1047951117002773

## Original Article

# Validation of the Brief Developmental Assessment in pre-school children with heart disease

Katherine L. Brown,<sup>1</sup> Deborah A. Ridout,<sup>2</sup> Christina Pagel,<sup>3</sup> Monica Lakhanpaul,<sup>4</sup> Suzan Kakat,<sup>1</sup> Victoria Banks,<sup>1</sup> Rodney Franklin,<sup>5</sup> Thomas Witter,<sup>6</sup> Rhian Lakhani,<sup>6</sup> Shane Tibby,<sup>6</sup> David Anderson,<sup>6</sup> Victor Tsang,<sup>1</sup> Aparna U. Hoskote,<sup>1</sup> Jo Wray<sup>1</sup>

<sup>1</sup>Charles West Division, NHS Foundation Trust, Great Ormond Street Hospital, Great Ormond Street; <sup>2</sup>Population, Policy and Practice Programme, UCL, Great Ormond Street Institute of Child Health; <sup>3</sup>Clinical Operational Research Unit, University College London; <sup>4</sup>Community Child Health, UCL, Great Ormond Street Institute of Child Health; <sup>5</sup>Paediatric Cardiology Department, Royal Brompton and Harefield NHS Foundation Trust; <sup>6</sup>Department Paediatric Cardiology and Cardiac Surgery, Evelina London Children's Hospital, London, United Kingdom

**Abstract** *Introduction:* The objective of this study was to prospectively validate the “Brief Developmental Assessment”, which is a new early recognition tool for neurodevelopmental abnormalities in children with heart disease that was developed for use by cardiac teams. *Methods:* This was a prospective validation study among a representative sample of 960 pre-school children with heart disease from three United Kingdom tertiary cardiac centres who were analysed grouped into five separate age bands. *Results:* The “Brief Developmental Assessment” was successfully validated in the older four age bands, but not in the youngest representing infants under the age of 4 months, as pre-set validation thresholds were met – lower 95% confidence limit for the correlation coefficient above 0.75 – in terms of agreement of scores between two raters and with an external measure the “Mullen Scales of Early Learning”. On the basis of American Association of Pediatrics Guidelines, which state that the sensitivity and specificity of a developmental screening tool should fall between 70 and 80%, “Brief Developmental Assessment” outcome of Red meets this threshold for detection of Mullen scores >2 standard deviations below the mean. *Conclusion:* The “Brief Developmental Assessment” may be used to improve the quality of assessment of children with heart disease. This will require a training package for users and a guide to action for abnormal results. Further research is needed to determine how best to deploy the “Brief Developmental Assessment” at different time points in children with heart disease and to determine the management strategy in infants younger than 4 months old.

Keywords: CHD; neurodevelopmental outcome; assessment tool; psychosocial/quality of life

Received: 20 May 2017; Accepted: 14 November 2017

**D**EVELOPMENTAL AND BEHAVIOURAL DIFFICULTIES are common in children with heart disease (CHD) and may affect intelligence,<sup>1,2</sup> academic achievement,<sup>3,4</sup> language (development, expressive, and receptive),<sup>4,5</sup> visual construction and

perception,<sup>6,7</sup> attention,<sup>8,9</sup> executive functioning,<sup>10</sup> fine motor skills,<sup>10</sup> gross motor skills<sup>11,12</sup>, and psychosocial maladjustment (internalising and externalising problems).<sup>13,14</sup> The American Association of Pediatrics 2012 statement on neurodevelopmental outcomes in children with CHD<sup>15</sup> concludes that over and above what is offered to healthy children, those with CHD require “periodic developmental surveillance, screening, evaluation, and re-evaluation throughout childhood may enhance identification of

Correspondence to: K. L. Brown, Charles West Division, Great Ormond Street Hospital NHS Foundation Trust, Great Ormond Street, London WC1N 3JH, United Kingdom. Tel: +44 207 8138180; Fax: +44 207 8298673; E-mail: Katherine.brown@gosh.nhs.uk

significant deficits, allowing for appropriate therapies and education to enhance later academic, behavioural, psychosocial, and adaptive functioning”.

Stakeholders, especially parent groups, view the evaluation of developmental and behavioural difficulties in children with CHD as one of their highest priorities.<sup>16</sup> Although the benefits of early intervention for developmental difficulties has not been specifically studied in children with CHD, early intervention has been recommended across a range of paediatric contexts<sup>17</sup> and is supported by studies of premature infants<sup>18</sup> and autism;<sup>19</sup> early recognition and intervention is also more satisfactory to families.<sup>20</sup>

As part of wider health screening in the United Kingdom, all children undergo periodic clinical checks with a community-based nurse referred to as a health visitor; these do not rely on formal developmental tests apart from the “Ages and Stages Questionnaire”, which is used with children at the age of 2 years. Despite the recognised higher risk of children with CHD for developmental delay, they do not routinely undergo any additional developmental scrutiny over and above what is offered to the wider population of children, nor do they undergo any standardised assessment of neurodevelopment either before or after cardiac operations. Health professionals trained in cardiac-related specialities – paediatric cardiologists, paediatric cardiac surgeons and paediatric cardiac nurses – are not also trained as specialists in paediatrics or in child development in the United Kingdom, and therefore do not consider themselves equipped to undertake such assessments. Paediatric cardiac services in the United Kingdom are centralised to 11 high-volume centres, which are responsible for overseeing or undertaking the majority of the follow-up of infants and young children with heart disease, presenting a potential opportunity for children to undergo periodic additional assessment of neurodevelopment. Resources are not available for every child with CHD to undergo full neurodevelopmental assessment by a child development specialist, and therefore an alternative approach to this clinical problem is necessary. The “Brief Developmental Assessment” was developed with a view to bridging this gap in care within the National Health Service, and the primary aim of this study was to prospectively validate the “Brief Developmental Assessment”.<sup>21</sup>

### The “Brief Developmental Assessment”

The “Brief Developmental Assessment”, in terms of its rationale, development, and wider qualities, is described in more detail in our companion paper (reference number 21). The “Brief Developmental Assessment” is an early recognition tool for childhood

neurodevelopment that contains both direct observation and parental report. The “Brief Developmental Assessment” can be used without additional special equipment that is commonly necessary for child developmental assessment. Nurses or doctors who do not have specific training in child development but who have been trained in its use may undertake the “Brief Developmental Assessment”. The “Brief Developmental Assessment” takes 5–10 minutes to complete and is designed for pre-school children up to the age of 5 years and consists of five individual age bands of 0–16.9 weeks, 17–34.9 weeks, 35–60 weeks, 15 months to 2.9 years, and 3.0–4.9 years. The “Brief Developmental Assessment” covers all the measurable areas of child development relevant to CHD in children under the age of 5 years, within the domains of gross motor skills, fine motor skills, daily living skills, communication, and socialisation, as well as general understanding for the oldest age band.

“Brief Developmental Assessment” scores of amber – possible abnormality based on age appropriate milestones – and red – likely abnormality based on age-appropriate milestones – have been defined in order that results generate a useful guide to action. Within each developmental domain, individual items are scored as “yes” or “no” based on the child’s undertaken activity. Within each age band, there are four subsections for scoring based on gradations of age, hence based on the child’s exact age. The score for each domain is graded as red, amber, or green. The overall “Brief Developmental Assessment” result is graded red if there are any red domains. The “Brief Developmental Assessment” result is graded amber if there are any amber domains but no red domains. If all domains are green the “Brief Developmental Assessment” result is green.

An example of the Brief Developmental Assessment version applicable to age group 17–34.9 weeks is shown for reader information in the supplementary material.

### Methods

#### *Study population*

The study setting was at the three tertiary children’s cardiac centres based in London and the time period of the study was January 2014 to July 2015. By approaching all children between birth and 5 years of age with heart disease attending outpatient or inpatient settings excluding those who were clinically unwell within the intensive care unit, a representative convenience sample of 200 patients with heart disease within each age band was recruited. Once the target of 200 participants was reached in any age band, recruitment to that age band ceased. Children were assessed by a small team of trained psychology

assistants under the supervision of a single senior psychological researcher and a medical lead.

### *Evaluations undertaken*

The internal consistency of the “Brief Developmental Assessment” was assessed for items within each domain, between totals for domains, and across the entire measure based on Cronbach’s  $\alpha$ .

The internal reliability of the “Brief Developmental Assessment” was evaluated in terms of inter-rater agreement when two raters simultaneously and independently performed and scored the “Brief Developmental Assessment”. The study team consisted of at least three raters throughout the study duration, and the inter-rater performance assessment was undertaken whenever two raters could be scheduled to be free at the same time; again, recruitment ceased as soon as the required number of patients was recruited.

External measures used for validation of the “Brief Developmental Assessment” were Mullen Scales of Early Learning (Mullen)<sup>22</sup> and the Ages and Stages Questionnaire (Ages and Stages),<sup>23</sup> the details of which are presented in Table 1.

Concurrent validity of “Brief Developmental Assessment” scores was assessed against Mullen scores.

Evaluation of construct validity was based on detection of known abnormalities, and to that end study participants were defined as falling within a “known group” where the child had been diagnosed to have a condition linked to neurodevelopmental problems – a congenital syndrome,<sup>24</sup> an acquired condition such as stroke,<sup>25</sup> or a previously diagnosed developmental delay based on a specialist clinical assessment even if the cause was not stated – and when the Mullen result was amber or red.

Construct validity was further assessed by calculating sensitivity and specificity of the “Brief Developmental Assessment” for detection of neurodevelopmental abnormalities against both of the external measures.

### *Data analysis*

For the evaluation of agreement of “Brief Developmental Assessment” scores between two independent raters and with Mullen, the following measures were taken; see the “Development of the ‘Brief Developmental Assessment’”<sup>21</sup> and Table 1 for further details of measures:

- Each of the five age bands was analysed individually.
- Raw scores were used in order to remove within-measure age standardisation, as this standardisation is undertaken differently within the “Brief Developmental Assessment” and the Mullen.

- Cognitive and gross motor scores were analysed separately because this reflects the validated scoring protocol for the Mullen, which contains a cognitive summary score covering the domains of visual reception, fine motor, receptive language, and expressive language combined and a fifth individual domain of gross motor function separated.
- Gross motor comparisons were available under the age of 33 months, only reflecting the protocol for the Mullen where this domain is only assessed in younger children aged less than this cut-off.

Inter-rater agreement was judged based on intra-class correlation coefficients for raw “Brief Developmental Assessment” general total scores that reflect the cognitive domains combined and weighed  $\kappa$  statistics for the ordinal measure raw “Brief Developmental Assessment” gross motor scores. Successful validation was defined as the lower 95% confidence limit for the intra-class correlation coefficient (or weighted  $\kappa$ ) exceeding 0.75.

For comparison with the Mullen, the raw “Brief Developmental Assessment” scores for the first 100 recruited patients were used to generate regression models for predicting the raw Mullen scores. These predictions were then tested in the subsequent 100 recruited patients in each age band for both “Brief Developmental Assessment” general scores and “Brief Developmental Assessment” gross motor raw scores. Successful validation was defined as the lower 95% confidence limit for the intra-class correlation coefficient (or weighted  $\kappa$ ) between observed and predicted Mullen scores exceeding 0.75 in the test sample.

Although the “Brief Developmental Assessment” represents an early recognition tool (rather than a developmental screener), the threshold for acceptability with respect to detection of known abnormalities and sensitivity/specificity was based on the American Academy of Pediatrics Committee on Children with Disabilities 2001 guidelines for desirable sensitivity and specificity for a developmental screening tool of 70–80%.<sup>26</sup>

### *Sample size calculations*

For the evaluation of inter-rater reliability of the “Brief Developmental Assessment”, we required 56 patients per age band to estimate an expected inter-rater intra-class correlation of 0.9 with a precision of 5% – lower bound of 95% CI is 0.85.

For the agreement of the “Brief Developmental Assessment” with the Mullen, we required 200 patients per age band to allow us to estimate an intra-class correlation coefficient of 0.8 with 5% precision – lower bound of 95% CI is 0.75.

Table 1. Overview of external measures used to assess “Brief Developmental Assessment” performance including definition of outcomes.

Measure	Pertinent features of the measure	Protocol for scoring of the measure as stated in published instructions	Outcome definitions used in the validation study
Primary assessment: The Mullen Scales of Early Learning	<p>This is a validated measure for early developmental assessment between birth and 5 years.<sup>32</sup> It has been used in infant CHD for developmental surveillance<sup>33</sup></p> <p>There are five individual scales: four cognitive scales of visual reception, fine motor, receptive language, and expressive language, and a fifth individual scale of gross motor function applicable from birth to 33 months</p> <p>Age range is equivalent to the study population of pre-school-age children</p> <p>It takes only 30–40 minutes to complete, thus increasing its acceptability to research participant families attending clinic</p>	<p>The “raw” scores for four cognitive scales and separately for the gross motor are computed to form age-standardised “T scores” in each area</p> <p>Mean “T scores” for each scale within the general population are 50 with standard deviation 10</p> <p>The cognitive “T scores” applicable to the four cognitive scales combined may be further computed to generate a composite score, which within the general population has a mean of 100 with standard deviation 15</p>	<p>Mullen standardised scores<sup>22</sup> categorised as follows: Patients with age-standardised cognitive or gross motor score falling between 1 and 2 standard deviations below the mean were scored as amber (<i>Cognitive scores 70–84, gross motor scores 30–39</i>)</p> <p>Patients with age-standardised cognitive or gross motor score more than 2 standard deviations below the mean were scored as red (<i>Cognitive scores &lt;70, gross motor scores &lt;30</i>)</p> <p>All other patients are classified as green</p>
Secondary assessment: Ages and Stages Questionnaire-3	<p>This is a validated screening questionnaire consisting of 21 age-versions applicable between birth and 66 months.<sup>34</sup> It has been used in infant CHD for surveillance<sup>35</sup></p> <p>There are five developmental domains with responses based on parental report: communication, gross motor, fine motor, problem-solving, and personal social</p> <p>It is used with the aim of capturing the Adaptive, and Social and Emotional domains not covered by the Mullen</p> <p>Of note, it does not entail direct observation</p>	<p>Each item is scored depending upon whether the child performs consistently (10 points), sometimes (5 points), or not yet (0 points)</p> <p>The total achievable score for each domain ranges from 0 to 60. On the basis of the published means and standard deviations for each age-version questionnaire, two thresholds have been established for each tested area to define a child’s score as “close to cut-off” (between 1 and 2 standard deviations from normative mean) and “below the normal range” (at least 2 standard deviations from the normative mean)</p>	<p>Parental responses within the questionnaires were categorised as follows, which is based on the manual for usage<sup>23</sup>:</p> <p>Patients with any domain falling between 1 and 2 standard deviations below the normative mean were scored as amber</p> <p>Patients with any domain scores more than 2 standard deviations below the normative mean were scored as red.</p> <p>All other patients are classified as green</p>

Within each of the age bands – excluding the youngest infants – a sample size of 200, comprising ~50 children with known developmental abnormality and 150 children presumed to have normal development, with anticipated prevalence of 25%<sup>27</sup>, provided sufficient numbers to detect a 0.5 SD difference in mean “Brief Developmental Assessment” scores between known groups, with 80% power and 5% significance. When assessing the ability of the “Brief Developmental Assessment” to discriminate between children with and without developmental abnormalities, the study was powered to detect a developmental abnormality with 12% precision, for an assumed sensitivity of 80%. We anticipated that the use of the “Brief Developmental Assessment” would result in a lower specificity, possibly 65%, and for this our sample provided 8% precision for this estimate. We were less concerned about the level of specificity as false positives where a child is subjected to medical review are unlikely to be harmful. Furthermore, we expected the “Brief Developmental Assessment” to have a higher sensitivity of 90% for detecting severe developmental abnormalities, and thus for a conservative estimate of prevalence for severe cases of 10% our sample size of 200 would provide a precision of 14%.

## Results

The case mix of 982 consented participants in the study is shown in Table 2. The age distribution of the sample is skewed towards younger infants because the width of the five age bands narrows as age falls; the median age across all age bands is 11.5 months – interquartile range, 5 months to 2.6 years.

### Internal reliability

The internal reliability of the “Brief Developmental Assessment”, expressed as Cronbach’s  $\alpha$ , is shown in Table 3. This was high between “Brief Developmental Assessment” total scores and between all items but low in selected domains of the “Brief Developmental Assessment” particularly within the youngest two age bands representing children under 8 months of age.

### Inter-rater reliability

A total of 160 children participated in the evaluation of inter-rater reliability of the “Brief Developmental Assessment” (see Table 4). Correlations were very high for all age bands, thus passing the pre-set threshold for inter-rater validity.

Table 2. Description of the demographics and clinical features of study cohort by age band.

	Age band one (0–16 weeks)	Age band two (17–34 weeks)	Age band three (35–60 weeks)	Age band four (15 months to 2.9 years)	Age band five (3.0–5 years)
Number of patients included	199	188	192	198	205
Number (%) of males	110 (55%)	82 (44%)	106 (55%)	103 (52%)	103 (50%)
Median age IQR (months)	2.2 (1.4, 3.1)	6.2 (5.1, 7.1)	11.4 (10.0, 13.2)	25.2 (20.8, 30.9)	48.0 (41.9, 53.8)
Number (%) with single-ventricle circulation	25 (13%)	12 (6%)	9 (5%)	23 (12%)	25 (12%)
Number (%) with more than 1 operation	16 (8%)	15 (8%)	15 (8%)	33 (17%)	51 (25%)
Number (%) with more than 1 catheter	4 (2%)	6 (3%)	7 (4%)	7 (4%)	11 (5%)
No known developmental delay or linked condition	163 (81.9%)	138 (74.2%)	147 (76.6%)	154 (77.8%)	149 (72.3%)
Congenital condition linked to developmental delay	23 (11.6%)	35 (18.8%)	39 (20.3%)	30 (15.2%)	34 (16.5%)
Developmental delay of unknown cause	4 (2%)	4 (2.1%)	4 (2.1%)	7 (3.5%)	15 (7.3%)
Acquired brain injury	9 (4.5%)	8 (4.3%)	1 (0.5%)	7 (3.5%)	6 (2.9%)
A combination of developmental delay-related diagnoses	0	1 (0.5%)	1 (0.5%)	0	2 (1%)

Table 3. Description of the internal reliability of the “Brief Developmental Assessment” by age band.

Internal reliability of based on Cronbach’s $\alpha$	Age band one (0–16 weeks)	Age band two (17–34 weeks)	Age band three (35–60 weeks)	Age band four (15 months to 2.9 years)	Age band five (3.0–5 years)
Number of patients included	199	188	192	198	205
Between domain totals	0.84	0.86	0.88	0.89	0.90
Between all items	0.84	0.86	0.90	0.91	0.92
Between items within gross motor domain	0.59	0.51	0.82	0.68	0.69
Within fine motor domain	0.46	0.71	0.49	0.61	0.70
Within daily living skills domain	0.40	0.60	0.58	0.55	0.67
Within communication domain	0.51	0.36	0.61	0.82	0.81
Within socialisation skills domain	0.52	0.33	0.61	0.57	0.51
Within cognition domain	NA	0.45	0.46	0.58	0.80

Table 4. Inter-rater agreement of the “Brief Developmental Assessment” and concurrent validity of the “Brief Developmental Assessment” against the Mullen expressed as correlation by age band.

Definition of success		Age band one (0–16 weeks)	Age band two (17–34 weeks)	Age band three (35–60 weeks)	Age band four (15 months to 2.9 years)	Age band five (3.0–5 years)
<b>Validation goal for “Brief Developmental Assessment”</b>						
Number of patients included		36	25	29	35	35
Inter-rater agreement cognitive scores	Intra-class correlation: lower 95% confidence interval >0.75	0.99 (0.98, 0.99)	0.99 (0.98, 1.00)	0.99 (0.99, 1.00)	1.00 (0.99, 1.00)	1.00 (1.00, 1.00)
Inter-rater agreement gross motor scores	Weighted $\kappa$ : lower 95% confidence interval >0.75	0.93 (0.76, 1.00)	0.98 (0.92, 1.00)	0.99 (0.96, 1.00)	1.00 (0.99, 1.00)	0.99 (0.98, 1.00)
<b>Comparison between “Brief Developmental Assessment” and Mullen</b>						
Number of patients in the test set		99	88	92	98	105
Agreement of raw cognitive scores (test data)	Intra-class correlation: lower 95% confidence interval >0.75	0.75 (0.67, 0.83)	0.86 (0.80, 0.91)	0.90 (0.85, 0.94)	0.91 (0.87, 0.94)	0.91 (0.88, 0.94)
Agreement of raw gross motor scores (test data)	Intra-class correlation: lower 95% confidence interval >0.75	0.71 (0.62, 0.79)	0.68 (0.58, 0.78)	0.84 (0.78, 0.89)	0.83 (0.78, 0.89)	0.83 (0.78, 0.89)

#### *Agreement between the “Brief Developmental Assessment” and the Mullen*

Of the 981 participants, 21 did not complete one or more domains of the Mullen, and thus a total of 960 children participated in the evaluation of concurrent validity of “Brief Developmental Assessment” against the Mullen (see Table 4). For age bands two to five, the pre-set thresholds were met with the exception of gross motor in age band two; however, in age band one the “Brief Developmental Assessment” displayed a much weaker correlation with the Mullen and therefore did not pass the pre-set threshold for validity.

#### *Developmental outcomes*

The developmental outcomes of participants based on the “Brief Developmental Assessment”, Mullen, and Ages and Stages are presented in Table 5 and summarised as follows:

There were 960 children completing both the “Brief Developmental Assessment” and the Mullen:

For “Brief Developmental Assessment”, 364 (38%) had a green result, 361 (38%) had an amber result, and 235 (24%) had a red result.

For Mullen, and considering both Mullen cognitive composite scores and, where applicable based on age, gross motor scores, 639 (67%) had a green result, 178 (18%) had an amber result, and 143 (15%) had a red result.

Data were missing for at least one Ages and Stages domain in 149 children (15%), and all of these children were excluded from validity analyses involving

the Ages and Stages. Of 832 children completing the Ages and Stages, only 238 (29%) had a green result, whereas 213 (25%) had an amber result and 381 (46%) had a red result.

#### *Construct validity*

Of the 960 participants who undertook the Mullen, 227 (24%) had a condition linked to developmental delay, of whom 153 (67%) also had a red or amber Mullen result, thus meeting the criteria for a “known group” with which to evaluate construct validity. Of these, 141 (92%) were also detected based on a red or amber “Brief Developmental Assessment” result, thus passing the pre-set threshold of 80%.

Surprisingly, 74 (33%) children with a condition linked to developmental delay had Mullen result of green. Of these 74, 40 (54%) were under the age of 8 months, and therefore based on young age developmental delay may not yet be evident. Furthermore, although 17 (23%) had a defined genetic condition such as Down’s syndrome, 13 (18%) had peri-operative neurological events of unknown significance and the remaining 44 (59%) had a range of congenital abnormalities where development incorporates a range of outcomes including normality.

Moreover, there were 168 children with a Mullen result of red or amber representing 18% of the study cohort who were not in a known group – that is, they had no known genetic or acquired condition linked to developmental problems and no known diagnosis of developmental delay stated by either the parents or

Table 5. Descriptive table of developmental scores and known groups by age band.

	Age band one (0–16 weeks)	Age band two (17–34 weeks)	Age band three (35–60 weeks)	Age band four (15 months to 2.9 years)	Age band five (3.0–5 years)
<b>Cognitive</b>					
Mullen Red	2 (1%)	6 (3%)	12 (6%)	32 (17%)	37 (19%)
Mullen Amber	24 (12%)	19 (10%)	21 (11%)	32 (17%)	18 (9%)
Mullen Green	171 (87%)	161 (87%)	157 (83%)	128 (66%)	140 (72%)
<b>Gross motor</b>					
Mullen Red	0	22 (12%)	40 (21%)	24 (14%)	N/A
Mullen Amber	25 (13%)	28 (15%)	47 (25%)	22 (13%)	N/A
Mullen Green	172 (87%)	136 (73%)	103 (54%)	123 (73%)	N/A
<b>Cognitive and gross motor combined result</b>					
Mullen Red	2 (1%)	24 (13%)	40 (21%)	40 (21%)	37 (19%)
Mullen Amber or Red	38 (19%)	57 (31%)	90 (47%)	81 (42%)	55 (28%)
					cognitive only
Ages and Stages Red	99 (62%)	82 (52%)	84 (49%)	63 (37%)	53 (31%)
Ages and Stages Amber	36 (22%)	40 (26%)	43 (25%)	56 (33%)	38 (22%)
Ages and Stages Green	26 (16%)	34 (22%)	45 (26%)	52 (30%)	81 (47%)
Brief Developmental Assessment Red	38 (19%)	49 (26%)	54 (28%)	61 (31%)	39 (19%)
Brief Developmental Assessment Amber	84 (42%)	77 (42%)	62 (32%)	75 (38%)	74 (36%)
Brief Developmental Assessment Green	77 (39%)	60 (32%)	76 (40%)	62 (31%)	93 (45%)
Known group – all types	36 (18%)	48 (26%)	45 (23%)	44 (22%)	57 (28%)
Known group + Mullen Red	1 (0.5%)	14 (8%)	24 (13%)	24 (13%)	32 (16%)
Known group + Mullen Red or Amber	13 (7%)	30 (16%)	35 (18%)	34 (18%)	41 (21%)

The number of patients completing the Brief Developmental Assessment and the Mullen was 960; and the Ages and Stages was 832.

written anywhere in their medical records that were based at the tertiary hospital. The charts of these patients were reviewed (see discussion section) and this finding may relate to under-detection of true abnormalities in the study population.

As might be expected given that child development assessments in general are more reliable in older children, the sensitivity and specificity of the “Brief Developmental Assessment” against the external measures improved with increasing age, with the best performance in age band five and poorest performance in age band one, which did not meet the criteria for validity. Given that the Ages and Stages is based on parental report only, whereas the Mullen is an objective validated developmental test, as expected, the “Brief Developmental Assessment” performed better against the Mullen than the Ages and Stages. The construct validity in age bands two to five combined may be summarized as follows (refer to (Table 6):

The test measure of “Brief Developmental Assessment” red or amber has excellent sensitivity against the Mullen and good sensitivity against the Ages and Stages, but moderate to low specificity for both external measures.

The test measure of “Brief Developmental Assessment” red has variable sensitivity but high specificity for both external measures. When considered based on American Association of Pediatrics Standards for the performance of a developmental screening tool, which state that the sensitivity and specificity of a developmental screening tool should fall between 70 and 80%<sup>26</sup>, the “Brief Developmental Assessment”

outcome of red against an outcome of Mullen red is compliant.

Positive and negative predictive values, as well as comparisons between the Ages and Stages and the Mullen, are presented for information purposes in Table 6.

## Discussion

### *Summary of validation*

The primary aim of this study, which was to validate the “Brief Developmental Assessment” as an early recognition tool for childhood developmental delay, was achieved within a population with heart disease between the ages of 4 months and 5 years. Previous researchers have presented sensitivity and specificity across a range of thresholds as a method to judge the performance of a new test against validated measures and have used this approach to select the optimal threshold to trigger an abnormal result<sup>28,29</sup> as has been undertaken in this study. These analyses support the use of both “Brief Developmental Assessment” results of amber and red as thresholds to trigger further evaluation of a child, although after reassessment a proportion of such children may not turn out to have developmental delay. The protocol for such reassessment is currently being delineated within a Delphi survey and goes beyond the scope of the current study. The Delphi survey entails a series of questions to a large group of health professionals from a range of settings and backgrounds that seeks to achieve a consensus as to the referral and

Table 6. Sensitivity and specificity of test measures in age bands two to five, combined.

Test measure	Comparison measure (measure utilised as gold standard for each specific comparison)	Sensitivity % (95% confidence interval)	Specificity % (95% confidence interval)	Positive predictive value %	Negative predictive value %	Abnormality based on test measure (%)	Abnormality based on comparison measure (%)
Brief Developmental Assessment	Ages and Stages						
Red or Amber	Red or Amber	70.2 (65.7, 74.3)	59.4 (52.5, 66.1)	78.9	47.9	408 (60.8)	459 (68.4)
Red or Amber	Red	78.7 (73.5, 83.4)	52.2 (47.1, 57.2)	54.4	77.2	408 (60.8)	282 (42.0)
Red	Red or Amber	32.7 (28.4, 37.2)	92.5 (88.0, 95.6)	90.4	38.8	166 (24.7)	459 (68.4)
Red	Red	42.6 (36.7, 48.6)	88.2 (84.5, 91.2)	72.3	67.9	166 (24.7)	282 (42.0)
	Mullen cognitive						
Red or Amber	Red or Amber	92.1 (87.1, 95.6)	46.6 (42.5, 50.7)	34.2	95.1	476 (62.4)	177 (23.2)
Red or Amber	Red	100.0 (95.8, 100)	42.5 (38.7, 46.3)	18.3	100.0	476 (62.4)	87 (11.4)
Red	Red or Amber	55.9 (48.3, 63.4)	83.3 (80.0, 86.2)	50.3	86.2	197 (25.8)	177 (23.2)
Red	Red	70.1 (59.4, 79.5)	79.9 (76.7, 82.8)	31.0	95.4	197 (25.8)	87 (11.4)
	Mullen cognitive and or gross motor						
Red or Amber	Red or Amber	85.9 (81.3, 89.7)	51.5 (46.9, 56.0)	51.1	86.1	476 (62.4)	283 (37.1)
Red or Amber	Red	99.3 (96.1, 100)	46.0 (42.0, 50.0)	29.4	99.7	476 (62.4)	141 (18.5)
Red	Red or Amber	48.1 (42.1, 54.0)	87.3 (84.0, 90.1)	69.0	74.0	197 (25.8)	283 (37.1)
Red	Red	70.2 (61.9, 77.6)	84.2 (81.1, 87.0)	50.3	92.6	197 (25.8)	141 (18.5)
Ages and Stages	Mullen cognitive						
Red or Amber	Red or Amber	96.6 (92.2, 98.9)	39.6 (35.3, 44.0)	31.4	97.6	449 (68.4)	146 (22.3)
Red or Amber	Red	98.5 (92.0, 100)	35.0 (31.1, 39.0)	14.7	99.5	449 (68.4)	67 (10.2)
Red	Red or Amber	81.5 (74.2, 87.4)	68.8 (64.6, 72.8)	42.8	92.9	278 (42.4)	146 (22.3)
Red	Red	95.5 (87.5, 99.1)	63.7 (59.6, 67.6)	23.0	99.2	278 (42.4)	67 (10.2)
	Mullen cognitive and or gross motor						
Red or Amber	Red or Amber	95.8 (92.5, 98.0)	47.4 (42.5, 52.3)	51.2	95.2	449 (68.4)	240 (36.6)
Red or Amber	Red	99.1 (95.3, 100)	38.1 (34.0, 42.3)	25.4	99.5	449 (68.4)	115 (17.5)
Red	Red or Amber	77.9 (72.1, 83.0)	78.1 (73.8, 82.0)	67.3	86.0	278 (42.4)	240 (36.6)
Red	Red	97.4 (92.6, 99.5)	69.3 (65.2, 73.2)	40.3	99.2	278 (42.4)	115 (17.5)

The number of patients completing the Brief Developmental Assessment and Mullen was 763 and the Ages and Stages was 671. Data from age band one not included.

reassessment pathway for children who have either amber or a red “Brief Developmental Assessment” test result picked up by the cardiac team at the tertiary centre.

#### Limitations of the validation

Our positive evaluation of “Brief Developmental Assessment” validity and reliability relates only to tests at a single time point and the constructs of test–retest validity and responsiveness over time could not be assessed within the scope of this study. Both concepts are challenging to test within a rapidly developing population of very young children with a significant health problem such as CHD and a further dedicated study will be required to explore these in particular repeated testing over time.

We note that the Brief Developmental Assessment has been developed as an early recognition tool for child development, and it is not intended to replace full formal neurodevelopment evaluation. It is our hope and intent that children flagged up by the Brief Developmental Assessment when it is used with them by the cardiac team will be speedily referred to and assessed by a neurodevelopmental clinic with a

more detailed formal evaluation using gold standard neurodevelopmental tests.

A motivation underpinning our study was a hypothesis that the processes in place to assess the neurodevelopment of children with CHD require improvement within the United Kingdom and children with CHD and developmental delay may be under-recognised. Indeed, 168 children with red or amber Mullen results were not in a known group, and chart review undertaken by one of three clinicians revealed concerns from the parents about the child’s development and/or other risk factors for abnormal development such as a history of cardiac arrest or mechanical circulatory support<sup>30,31</sup> in the majority. A review of the services that children were under and what actions might need to be taken to meet their needs is underway and goes beyond the scope of the validation study. Health professionals and parents have commented anecdotally that children with CHD, such as those under the age of 5 years recruited to this study, are in general undergoing treatments for their heart including surgery, and this represents the main focus of contacts with health professionals including both those at the cardiac centres and also in the community such as health visitors. This may be

represent a reason for these 168 children with red or amber Mullen results not already being identified as in a known group.

The Brief Developmental Assessment was developed for use with children who have heart disease, and has not been used or validated with healthy children. There is the potential that the Brief Developmental Assessment might be useful within other groups of children who for medical reasons are at greater risk of neurodevelopmental problems, such as survivors of other types of critical illness. However, in order to take this forwards, further testing and research would be required.

#### *Comment on external measures*

Parents who were concerned about their child's development preferred to watch the entirety of the testing with "Brief Developmental Assessment" and Mullen, and were less likely than parents who had no concerns about their child to complete the Ages and Stages while their child was being assessed. This is supported by a comparison of the Mullen cognitive results between children with missing Ages and Stages – 29% Mullen cognitive results red or amber – and those who completed Ages and Stages – 19% Mullen cognitive results red or amber. The overall proportion of red results for the Ages and Stages (46%) was very high, and perhaps proportion of Ages and Stages results that were red would have been even higher had the missing 15.2% been included.

This is not a cohort study with longitudinal follow-up thus limiting interpretation, but as displayed in Table 6 developmental delay based on Mullen results was detected least frequently in the youngest infants in contrast with developmental delays based on Ages and Stages results, which was detected most frequently in the youngest infants. Medical ill health in children with CHD is more prevalent in infancy as this is a period when interventions are commonly undertaken. These observations support a hypothesis that in children with CHD the Ages and Stages may be picking up a range of issues including developmental delay and general ill health, and further emphasises the importance of an initial evaluation for signs of developmental delay that incorporates direct observation of children with CHD, such as the "Brief Developmental Assessment" provides. Furthermore, this emphasises the recognised importance of periodic assessment of neurodevelopment over time in children with heart disease.

#### *Summary and next steps*

The development and validation of the "Brief Developmental Assessment" represents an

opportunity to improve the future quality of peri-intervention assessment for children with heart disease between the ages of 4 months and 5 years in United Kingdom children's cardiac centres. This initiative of using the "Brief Developmental Assessment" within the cardiac centres would complement the health visitor assessments that all children receive and will be undertaken by cardiac staff aware of details of the child's history such as cardiac arrests and mechanical circulatory support that predispose to neurodevelopmental problems.

One problem with the current system of surveillance for young children in the United Kingdom as it pertains to children with heart disease is that in addition to them being inherently at higher risk than other children and therefore potentially benefiting from additional scrutiny, the standard health visitor reviews correspond with a period in these children's lives when cardiac conditions are often having a significant impact, and this may be a barrier to the systems effectiveness for them, and may account in part for the occurrence of undetected neurodevelopmental problems in the population that we observed.

Roll out of the "Brief Developmental Assessment" will require a training package for users and a guide to action for abnormal results, such as a standardised report for specific relevant health professionals and parents. Since the "Brief Developmental Assessment" is a short assessment that is undertaken by staff with the training background of those working in a cardiac centre with whom children who have CHD are in regular contact early in life and without additional equipment successful implementation is more likely. Further research is needed to delineate the optimal approach to assessment of children over time including when to incorporate the "Brief Developmental Assessment" and to establish the most effective management strategy for infants who attend cardiac centres for interventions when they are younger than 4 months old as the "Brief Developmental Assessment" is not appropriate for them.

#### **Acknowledgements**

The authors would like to acknowledge the contributions of the following staff based in the paediatric cardiac units at Great Ormond Street Hospital, Evelina London, Royal Brompton Children's Hospital: research assistants Seona Granville, Anna Lucock, Jessica Davies, Jordana Walsh, Merari Ferreira, Sophie Good, Rachel Pulham, Jana Gurashvilli, and Mark Mayes; the data managers Sarah Bohanan and Abi Howarth; the nurses Angie Scarisbroke, Liz Smith, and

Teresa Johnson; and the doctors Owen Miller, Dora Wood, and Carol Bodlani. The authors would also like to acknowledge the invaluable advice of Martin Utley (London), Kate Bull (Cambridge), William Gaynor (Philadelphia), Lisa Allera (London), and the doctors based at the Community Paediatric Clinic in Harringey, North London.

Authors' Contributions: Conceptualisation and design of the study: K. Brown, D. Ridout, C. Pagel, M. Lakhanpaul, A. Hoskote, and J. Wray; supervision and day-to-day running of the study: S. Tibby, D. Anderson, R. Lakhani, T. Witter, R. Franklin, A. Hoskote, and J. Wray; analyses of study data: K. Brown, D. Ridout, C. Pagel, M. Lakhanpaul, S. Kakat, V. Banks, T. Witter, A. Hoskote, and J. Wray; write-up: K. Brown and D. Ridout; editing and approval of final manuscript: C. Pagel, M. Lakhanpaul, A. Hoskote, J. Wray, V. Tsang, S. Kakat, V. Banks, S. Tibby, D. Anderson, R. Lakhani, T. Witter, and R. Franklin.

### Financial Support

This project was funded by the National Institute for Health Research Health Services and Delivery Research programme (Project No: 12/5005/06). The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the NIHR HS&DR programme or the Department of Health. K. Brown, D. Ridout, A. Hoskote, J. Wray, and V. Tsang were supported by the National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children NHS Foundation Trust and University College London. M. Lakhanpaul was supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) North Thames at Bart's Health NHS Trust.

### Conflicts of Interest

None.

### Ethical Standards

London City Road Research Ethics Committee (REC) granted approval for the study number 14-LO-1442 in November 2013.

### Supplementary material

To view supplementary material for this article, please visit <https://doi.org/10.1017/S1047951117002773>

### References

- Bellinger DC, Wypij D, duPlessis AJ, et al. Neurodevelopmental status at eight years in children with dextro-transposition of the great arteries: the Boston Circulatory Arrest Trial. *J Thorac Cardiovasc Surg* 2003; 126: 1385–1396.
- Bellinger DC, Wypij D, Kuban KC, et al. Developmental and neurological status of children at 4 years of age after heart surgery with hypothermic circulatory arrest or low-flow cardiopulmonary bypass. *Circulation* 1999; 100: 526–532.
- Wernovsky G, Stiles KM, Gauvreau K, et al. Cognitive development after the Fontan operation. *Circulation* 2000; 102: 883–889.
- Mahle WT, Clancy RR, Moss EM, Gerdes M, Jobes DR, Wernovsky G. Neurodevelopmental outcome and lifestyle assessment in school-aged and adolescent children with hypoplastic left heart syndrome. *Pediatrics* 2000; 105: 1082–1089.
- Hovels-Gurich HH, Konrad K, Skorzewski D, et al. Long-term neurodevelopmental outcome and exercise capacity after corrective surgery for tetralogy of Fallot or ventricular septal defect in infancy. *Ann Thorac Surg* 2006; 81: 958–966.
- Hoffman GM, Mussatto KA, Brosig CL, et al. Systemic venous oxygen saturation after the Norwood procedure and childhood neurodevelopmental outcome. *J Thorac Cardiovasc Surg* 2005; 130: 1094–1100.
- Brosig CL, Mussatto KA, Kuhn EM, Tweddell JS. Neurodevelopmental outcome in preschool survivors of complex congenital heart disease: implications for clinical practice. *J Pediatr Health Care* 2007; 21: 3–12.
- Hovels-Gurich HH, Konrad K, Skorzewski D, Herpertz-Dahlmann B, Messmer BJ, Seghaye MC. Attentional dysfunction in children after corrective cardiac surgery in infancy. *Ann Thorac Surg* 2007; 83: 1425–1430.
- Shillingford AJ, Glanzman MM, Ittenbach RF, Clancy RR, Gaynor JW, Wernovsky G. Inattention, hyperactivity, and school performance in a population of school-age children with complex congenital heart disease. *Pediatrics* 2008; 121: e759–e767.
- Bellinger DC, Bernstein JH, Kirkwood MW, Rappaport LA, Newburger JW. Visual-spatial skills in children after open-heart surgery. *J Dev Behav Pediatr* 2003; 24: 169–179.
- Goldberg CS, Schwartz EM, Brunberg JA, et al. Neurodevelopmental outcome of patients after the fontan operation: a comparison between children with hypoplastic left heart syndrome and other functional single ventricle lesions. *J Pediatr* 2000; 137: 646–652.
- Williams DL, Gelijs AC, Moskowitz AJ, et al. Hypoplastic left heart syndrome: valuing the survival. *J Thorac Cardiovasc Surg* 2000; 119: 720–731.
- Miatton M, De Wolf D, Francois K, Thiery E, Vingerhoets G. Behavior and self-perception in children with a surgically corrected congenital heart disease. *J Dev Behav Pediatr* 2007; 28: 294–301.
- Hovels-Gurich HH, Konrad K, Wiesner M, et al. Long term behavioural outcome after neonatal arterial switch operation for transposition of the great arteries. *Arch Dis Child* 2002; 87: 506–510.
- Marino BS, Lipkin PH, Newburger JW, et al. Neurodevelopmental outcomes in children with congenital heart disease: evaluation and management: a scientific statement from the American Heart Association. *Circulation* 2012; 126: 1143–1172.
- Brown KL, Pagel C, Brimmell R, et al. Definition of important early morbidities related to paediatric cardiac surgery. *Cardiol Young* 2016; 1–10.
- Newman T, McEwen J, Mackin H, Slowley M. Improving the wellbeing of disabled children and their families through increasing the quality and range of early years interventions. In: Unit CEBPaR (ed.) *Disability Research Review*. Barnados, London, 2008: 1–123.
- Blauw-Hospers CH, Hadders-Algra M. A systematic review of the effects of early intervention on motor development. *Dev Med Child Neurol* 2005; 47: 421–432.

19. Zuckerman K, Lindly OJ, Chavez AE. Timeliness of autism spectrum disorder diagnosis and use of services among U.S. elementary school-aged children. *Psychiatr Serv* 2017; 68: 33–40.
20. Bailey DB Jr., Hebbeler K, Spiker D, Scarborough A, Mallik S, Nelson L. Thirty-six-month outcomes for families of children who have disabilities and participated in early intervention. *Pediatrics* 2005; 116: 1346–1352.
21. Wray J, Ridout D, Lakhanpaul M, et al. Development and preliminary testing of the Brief Development (BDA) – an early recognition tool for children with heart disease 2018; doi: 10.1017/S1047951117002918.
22. Mullen E. *Mullen Scales of Early Learning*. Pearson, USA, 1995.
23. Squires J. *The Ages and Stages Questionnaire 3rd edn*. Brookes Publishing, Baltimore, USA, 2006.
24. Wellesley D, Boyd P, Dolk H, Pattenden S. An aetiological classification of birth defects for epidemiological research. *J Med Genet* 2005; 42: 54–57.
25. Mahle WT, Tavani F, Zimmerman RA, et al. An MRI study of neurological injury before and after congenital heart surgery. *Circulation* 2002; 106: I109–I114.
26. Pediatrics AAo. Developmental surveillance and screening of infants and young children. *Pediatrics* 2001; 108: 421–432.
27. Knowles R. Modelling survival in children with serious congenital heart defects. In: PHD UoL (ed.) PHD thesis. University of London, London UK, 2010.
28. Steenis LJ, Verhoeven M, Hessen DJ, van Baar AL. Parental and professional assessment of early child development: the ASQ-3 and the Bayley-III-NL. *Early Hum Dev* 2015; 91: 217–225.
29. Limbos MM, Joyce DP. Comparison of the ASQ and PEDS in screening for developmental delay in children presenting for primary care. *J Dev Behav Pediatr* 2011; 32: 499–511.
30. Lequier L, Joffe AR, Robertson CM, et al. Two-year survival, mental, and motor outcomes after cardiac extracorporeal life support at less than five years of age. *J Thorac Cardiovasc Surg* 2008; 136: 976–983 e973.
31. Hamrick SE, Miller SP, Leonard C, et al. Trends in severe brain injury and neurodevelopmental outcome in premature newborn infants: the role of cystic periventricular leukomalacia. *J Pediatr* 2004; 145: 593–599.
32. Bishop SL, Guthrie W, Coffing M, Lord C. Convergent validity of the Mullen Scales of –and the differential ability scales in children with autism spectrum disorders. *Am J Intellect Dev Disabil* 2011; 116: 331–343.
33. Ungerleider RM, Shen I, Yeh T, et al. Routine mechanical ventricular assist following the Norwood procedure – improved neurologic outcome and excellent hospital survival. *Ann Thorac Surg* 2004; 77: 18–22.
34. Veldhuizen S, Clinton J, Rodriguez C, Wade TJ, Cairney J. Concurrent validity of the ages and stages questionnaires and Bayley developmental scales in a general population sample. *Acad Pediatr* 2015; 15: 231–237.
35. Goldberg CS, Lu M, Sleeper LA, et al. Factors associated with neurodevelopment for children with single ventricle lesions. *J Pediatr* 2014; 165: 490–496 e498.