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

Katherine L. Brown, Deborah A. Ridout, Christina Pagel, Monica Lakanpaul, Suzan Kakat, Victoria Banks, Rodney Franklin, Thomas Witter, Rhian Lakhani, Shane Tibby, David Anderson, Victor Tsang, Aparna U. Hoskote, Jo Wray

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
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. 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 and is supported by studies of premature infants and autism; early recognition and intervention is also more satisfactory to families.

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 and behavioural assessments. 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 α. 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) and the Ages and Stages Questionnaire (Ages and Stages), 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, an acquired condition such as stroke, 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’” 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 κ 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 κ) 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 κ) 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%.

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.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Pertinent features of the measure</th>
<th>Protocol for scoring of the measure as stated in published instructions</th>
<th>Outcome definitions used in the validation study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary assessment:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Mullen Scales of Early Learning</td>
<td>This is a validated measure for early developmental assessment between birth and 5 years. It has been used in infant CHD for developmental surveillance. 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. Age range is equivalent to the study population of pre-school-age children. It takes only 30–40 minutes to complete, thus increasing its acceptability to research participant families attending clinic.</td>
<td>The “raw” scores for four cognitive scales and separately for the gross motor are computed to form age-standardised “T scores” in each area. Mean “T scores” for each scale within the general population are 50 with standard deviation 10. 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.</td>
<td>Mullen standardised scores 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 (Cognitive scores &lt;70, gross motor scores &lt;30) Patients with age-standardised cognitive or gross motor score more than 2 standard deviations below the mean were scored as red (Cognitive scores &lt;70, gross motor scores &lt;30) All other patients are classified as green.</td>
</tr>
<tr>
<td><strong>Secondary assessment:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ages and Stages Questionnaire-3</td>
<td>This is a validated screening questionnaire consisting of 21 age-versions applicable between birth and 66 months. It has been used in infant CHD for surveillance. There are five developmental domains with responses based on parental report: communication, gross motor, fine motor, problem-solving, and personal social. It is used with the aim of capturing the Adaptive, and Social and Emotional domains not covered by the Mullen. Of note, it does not entail direct observation.</td>
<td>Each item is scored depending upon whether the child performs consistently (10 points), sometimes (5 points), or not yet (0 points). 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).</td>
<td>Parental responses within the questionnaires were categorised as follows, which is based on the manual for usage: Patients with any domain falling between 1 and 2 standard deviations below the normative mean were scored as amber. Patients with any domain scores more than 2 standard deviations below the normative mean were scored as red. All other patients are classified as green.</td>
</tr>
</tbody>
</table>
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%27, 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 α, 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.**

<table>
<thead>
<tr>
<th>Age band</th>
<th>Number of patients included</th>
<th>Number (%) of males</th>
<th>Median age IQR (months)</th>
<th>Number (%) with single-ventricle circulation</th>
<th>Number (%) with more than 1 catheter</th>
<th>No known developmental delay or linked condition</th>
<th>Congenital condition linked to developmental delay</th>
<th>Developmental delay of unknown cause</th>
<th>Acquired brain injury</th>
<th>A combination of developmental delay-related diagnoses</th>
</tr>
</thead>
<tbody>
<tr>
<td>One (0–16 weeks)</td>
<td>199</td>
<td>55%</td>
<td>1.4 (0.9, 2.1)</td>
<td>25%</td>
<td>4%</td>
<td>165 (81.9%)</td>
<td>33 (11.6%)</td>
<td>4 (2%)</td>
<td>9 (4.5%)</td>
<td>0</td>
</tr>
<tr>
<td>Two (17–34 weeks)</td>
<td>188</td>
<td>44%</td>
<td>5.1 (4.1, 6.1)</td>
<td>12%</td>
<td>6%</td>
<td>138 (74.2%)</td>
<td>35 (18.8%)</td>
<td>2 (1.1%)</td>
<td>8 (4.5%)</td>
<td>1 (0.5%)</td>
</tr>
<tr>
<td>Three (35–60 weeks)</td>
<td>192</td>
<td>55%</td>
<td>10.0 (8.8, 12.2)</td>
<td>9%</td>
<td>4%</td>
<td>147 (76.6%)</td>
<td>39 (20.3%)</td>
<td>4 (2.1%)</td>
<td>1 (0.5%)</td>
<td>0</td>
</tr>
<tr>
<td>Four (15 months to 2.9 years)</td>
<td>198</td>
<td>52%</td>
<td>13.2 (11.0, 15.2)</td>
<td>5%</td>
<td>3%</td>
<td>154 (77.8%)</td>
<td>30 (15.2%)</td>
<td>7 (3.5%)</td>
<td>1 (0.5%)</td>
<td>0</td>
</tr>
<tr>
<td>Five (3.0–5 years)</td>
<td>205</td>
<td>50%</td>
<td>20.9 (19.0, 22.9)</td>
<td>12%</td>
<td>5%</td>
<td>149 (72.3%)</td>
<td>34 (16.5%)</td>
<td>15 (7.3%)</td>
<td>6 (2.9%)</td>
<td>2 (1%)</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Internal reliability of based on Cronbach’s α</th>
<th>Age band one (0–16 weeks)</th>
<th>Age band two (17–34 weeks)</th>
<th>Age band three (35–60 weeks)</th>
<th>Age band four (15 months to 2.9 years)</th>
<th>Age band five (3.0–5 years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients included</td>
<td>199</td>
<td>188</td>
<td>192</td>
<td>198</td>
<td>205</td>
</tr>
<tr>
<td>Between domain totals</td>
<td>0.84</td>
<td>0.86</td>
<td>0.88</td>
<td>0.89</td>
<td>0.90</td>
</tr>
<tr>
<td>Between all items</td>
<td>0.84</td>
<td>0.86</td>
<td>0.90</td>
<td>0.91</td>
<td>0.92</td>
</tr>
<tr>
<td>Between items within gross motor domain</td>
<td>0.59</td>
<td>0.51</td>
<td>0.82</td>
<td>0.68</td>
<td>0.69</td>
</tr>
<tr>
<td>Within fine motor domain</td>
<td>0.46</td>
<td>0.71</td>
<td>0.49</td>
<td>0.61</td>
<td>0.70</td>
</tr>
<tr>
<td>Within daily living skills domain</td>
<td>0.40</td>
<td>0.60</td>
<td>0.58</td>
<td>0.55</td>
<td>0.67</td>
</tr>
<tr>
<td>Within communication domain</td>
<td>0.51</td>
<td>0.36</td>
<td>0.61</td>
<td>0.82</td>
<td>0.81</td>
</tr>
<tr>
<td>Within socialisation skills domain</td>
<td>0.52</td>
<td>0.33</td>
<td>0.61</td>
<td>0.57</td>
<td>0.51</td>
</tr>
<tr>
<td>Within cognition domain</td>
<td>0.56</td>
<td>0.45</td>
<td>0.46</td>
<td>0.58</td>
<td>0.80</td>
</tr>
</tbody>
</table>
Comparison between Validation goal for the gross motor in age band two; however, in age band one pre-set thresholds were met with the exception of fore did not pass the pre-set threshold for validity.

much weaker correlation with the Mullen and there-

“marised as follows:

Ages and Stages are presented in Table 5 and sum-

The developmental outcomes of participants based on Developmental outcomes

of developmental delay stated by either the parents or

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 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.

<table>
<thead>
<tr>
<th>Validation goal for “Brief Developmental Assessment”</th>
<th>Age band one (0–16 weeks)</th>
<th>Age band two (17–34 weeks)</th>
<th>Age band three (35–60 weeks)</th>
<th>Age band four (15 months to 2.9 years)</th>
<th>Age band five (3.0–5 years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients included</td>
<td>36</td>
<td>25</td>
<td>29</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>Inter-rater agreement cognitive scores</td>
<td>Intra-class correlation: lower 95% confidence interval &gt;0.75</td>
<td>0.99 (0.98, 0.99)</td>
<td>0.99 (0.98, 1.00)</td>
<td>0.99 (0.99, 1.00)</td>
<td>1.00 (0.99, 1.00)</td>
</tr>
<tr>
<td>Inter-rater agreement gross motor scores</td>
<td>Weighted κ: lower 95% confidence interval &gt;0.75</td>
<td>0.68 (0.67, 0.73)</td>
<td>0.93 (0.89, 0.96)</td>
<td>0.99 (0.98, 1.00)</td>
<td>1.00 (1.00, 1.00)</td>
</tr>
<tr>
<td>Comparison between “Brief Developmental Assessment” and Mullen</td>
<td>Number of patients in the test set</td>
<td>99</td>
<td>88</td>
<td>92</td>
<td>98</td>
</tr>
<tr>
<td>Agreement of raw cognitive scores (test data)</td>
<td>Intra-class correlation: lower 95% confidence interval &gt;0.75</td>
<td>0.75 (0.67, 0.83)</td>
<td>0.86 (0.80, 0.91)</td>
<td>0.90 (0.85, 0.94)</td>
<td>0.91 (0.87, 0.94)</td>
</tr>
<tr>
<td>Agreement of raw gross motor scores (test data)</td>
<td>Intra-class correlation: lower 95% confidence interval &gt;0.75</td>
<td>0.71 (0.62, 0.79)</td>
<td>0.68 (0.58, 0.78)</td>
<td>0.84 (0.78, 0.89)</td>
<td>0.83 (0.78, 0.89)</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Table 5. Descriptive table of developmental scores and known groups by age band.</th>
<th>Age band one (0–16 weeks)</th>
<th>Age band two (17–34 weeks)</th>
<th>Age band three (35–60 weeks)</th>
<th>Age band four (15 months to 2.9 years)</th>
<th>Age band five (3.0–5 years)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cognitive</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mullen Red</td>
<td>2 (1%)</td>
<td>6 (3%)</td>
<td>12 (6%)</td>
<td>32 (17%)</td>
<td>37 (19%)</td>
</tr>
<tr>
<td>Mullen Amber</td>
<td>24 (12%)</td>
<td>19 (10%)</td>
<td>21 (11%)</td>
<td>32 (17%)</td>
<td>18 (9%)</td>
</tr>
<tr>
<td>Mullen Green</td>
<td>171 (87%)</td>
<td>161 (87%)</td>
<td>157 (85%)</td>
<td>128 (66%)</td>
<td>140 (72%)</td>
</tr>
<tr>
<td><strong>Gross motor</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mullen Red</td>
<td>0</td>
<td>22 (12%)</td>
<td>40 (21%)</td>
<td>24 (14%)</td>
<td>N/A</td>
</tr>
<tr>
<td>Mullen Amber</td>
<td>25 (13%)</td>
<td>28 (15%)</td>
<td>47 (25%)</td>
<td>22 (13%)</td>
<td>N/A</td>
</tr>
<tr>
<td>Mullen Green</td>
<td>172 (87%)</td>
<td>136 (73%)</td>
<td>103 (54%)</td>
<td>123 (73%)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Cognitive and gross motor combined result</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mullen Red</td>
<td>2 (1%)</td>
<td>24 (13%)</td>
<td>40 (21%)</td>
<td>40 (21%)</td>
<td>37 (19%)</td>
</tr>
<tr>
<td>Mullen Amber or Red</td>
<td>38 (19%)</td>
<td>57 (31%)</td>
<td>90 (47%)</td>
<td>81 (42%)</td>
<td>55 (28%)</td>
</tr>
<tr>
<td>Ages and Stages Red</td>
<td>99 (62%)</td>
<td>82 (52%)</td>
<td>84 (49%)</td>
<td>63 (37%)</td>
<td>53 (31%)</td>
</tr>
<tr>
<td>Ages and Stages Amber</td>
<td>56 (22%)</td>
<td>40 (26%)</td>
<td>43 (25%)</td>
<td>56 (33%)</td>
<td>38 (22%)</td>
</tr>
<tr>
<td>Known group – all types</td>
<td>36 (16%)</td>
<td>34 (22%)</td>
<td>45 (26%)</td>
<td>52 (30%)</td>
<td>81 (47%)</td>
</tr>
<tr>
<td>Known group + Mullen Red</td>
<td>77 (39%)</td>
<td>49 (26%)</td>
<td>54 (28%)</td>
<td>61 (31%)</td>
<td>39 (19%)</td>
</tr>
<tr>
<td>Known group + Mullen Red or Amber</td>
<td>13 (7%)</td>
<td>57 (31%)</td>
<td>62 (32%)</td>
<td>75 (38%)</td>
<td>74 (36%)</td>
</tr>
<tr>
<td>Known group + Mullen Green</td>
<td>56 (28%)</td>
<td>48 (26%)</td>
<td>45 (25%)</td>
<td>44 (22%)</td>
<td>57 (28%)</td>
</tr>
<tr>
<td>Known group + Mullen Green or Amber</td>
<td>13 (7%)</td>
<td>30 (16%)</td>
<td>35 (18%)</td>
<td>34 (18%)</td>
<td>41 (21%)</td>
</tr>
</tbody>
</table>
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 neurodevelopmental 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 support30,31 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.


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. Lakanhpaul was supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) North Thames at Barr’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