

CLINICAL RESEARCH ARTICLE OPEN


# ASQ-3 and BSID-III's concurrent validity and predictive ability of cognitive outcome at 5 years

 Cian Duggan<sup>1✉</sup>, Alan D. Irvine<sup>2,3</sup>, Jonathan O'B Hourihane<sup>1,4</sup>, Mairead E. Kiely<sup>4,5</sup> and Deirdre M. Murray<sup>1,4</sup>

© The Author(s) 2023

**BACKGROUND:** Early detection of cognitive disability is challenging. We assessed the domain-specific, concurrent validity of the ages and stages questionnaire (ASQ-3) and the Bayley Scales of Infant and Toddler Development (BSID-III), and their ability to predict cognitive delay at school age.

**METHODS:** Within a longitudinal birth cohort study, a nested cohort of children was assessed using ASQ-3 and BSID-III at 24 months, and at 5 years using the Kaufmann brief IQ test (KBIT).

**RESULTS:** 278 children were assessed using BSID-III and ASQ-3 at 24-months; mean(SD) BW = 3445(506) grams, M:F ratio=52:48. ASQ-3 had reasonable predictive ability (AUROC, *p* value, sensitivity:specificity) of same domain delay for motor (0.630, *p* = 0.008, 50%:76.1%) and language (0.623, *p* = 0.010, 25%:99.5%) at 2 years, but poor ability to detect cognitive delay compared to BSID-III (0.587, *p* = 0.124, 20.7%/96.8%). 204/278 children were assessed at 5 years. BSID-III language and cognition domains showed better correlation with verbal and nonverbal IQ (*R* = 0.435, *p* < 0.001 and 0.388, *p* < 0.001 respectively). Both assessments showed high specificity and low sensitivity for predicting delay at 5 years.

**CONCLUSIONS:** The ASQ-3 cognitive domain showed poor concurrent validity with BSID-III cognitive score. Both ASQ-3 and BSID-III at 2 years poorly predict cognitive delay at 5 years.

*Pediatric Research* (2023) 94:1465–1471; <https://doi.org/10.1038/s41390-023-02528-y>

## IMPACT:

- The ASQ-3 does not adequately detect cognitive delay or predict cognitive delay at 5 years, particularly for children with mild to moderate delay.
- The ASQ-3 shows reasonable concurrent validity with the motor and language subscales of the BSID-III. Neither early screening nor formal developmental testing demonstrated significant predictive validity to screen for cognitive delay at school age.
- This article highlights the need to analyse our existing model of using the ASQ-3 to screen for cognitive delay in children aged 2 years.

## INTRODUCTION

Early intellectual function is a key predictor of adult health and well-being.<sup>1</sup> Pre-school identification of those most at risk allows for early intervention which improves developmental outcomes as demonstrated in a Cochrane review (2015) and by refs.<sup>2,3</sup> Current assessment methods rely on parental report or structured direct assessments based on the ascertainment of developmental milestones. These developmental milestones are, in themselves, approximate metrics for later cognition. The ICD-11 defines neurodevelopmental disorders as “*behavioural and cognitive disorders that arise during the developmental period that involve significant difficulties in the acquisition and execution of specific intellectual, motor, or social functions*”.<sup>4,5</sup> The prevalence of neurodevelopmental delay has been reported as high as fifteen to eighteen percent, with approximately three percent having severe delay.<sup>6–8</sup> Incidence increases significantly in high risk

groups such as small for gestational age<sup>9</sup> and preterm infants ranging from 50.2–62.5% in those born <29 weeks' gestation.<sup>10,11</sup> The American Academy of Paediatrics recommends periodic screening using a validated tool at nine, eighteen and thirty months of age to allow targeted early intervention.<sup>12</sup>

The Ages and Stages Questionnaire 3 (ASQ-3) is a parentally completed developmental screening tool which can be performed in 10–15 minutes<sup>5</sup> helping to identify children requiring further developmental evaluation.<sup>8</sup> The Bayley Scales of Infant and Toddler Development III (BSID-III) is a more comprehensive developmental assessment undertaken by a trained healthcare professional. These tools are further described in the methods section. Together, these tools comprise a valuable resource in clinical practice, however, there is a paucity of research, with often contradictory views on their validity, agreement and predictive value for later cognitive outcome. One systematic review identified three papers which

<sup>1</sup>Department of Paediatrics and Child Health, Cork University Hospital, Cork, Ireland. <sup>2</sup>Department of Clinical Medicine, Trinity College, Dublin, Ireland. <sup>3</sup>Department of Paediatric Dermatology, Our Lady's Children's Hospital, Dublin, Ireland. <sup>4</sup>The INFANT Research Centre, University College Cork, Cork, Ireland. <sup>5</sup>University College Cork, Cork Centre for Vitamin D and Nutrition Research, School of Food and Nutritional Sciences, Cork, Ireland. ✉email: [cianduggan25@gmail.com](mailto:cianduggan25@gmail.com)

Received: 27 July 2022 Revised: 20 January 2023 Accepted: 28 January 2023  
 Published online: 25 February 2023

assessed the ASQ-3's validity in conjunction with the BSID-III (the accepted gold standard at this age).<sup>13–16</sup> The ASQ-3 self-reports a sensitivity of 86% and a specificity of 85%.<sup>5</sup> However, independent studies report sensitivity and specificity ranging from 67–100% and 65–93%, respectively<sup>17–21</sup> when compared to the BSID-III (as a gold standard). Velikonja et al. indicated that the lack of research and quality of studies made it difficult to draw any clear conclusions on sensitivities nor specificities of the ASQ-3.<sup>16</sup> The ASQ has been demonstrated to have a higher sensitivity in high-risk groups, particularly for those with severe delay. Unfortunately, the rate of detection is markedly lower for mild and moderate delay, a group with a much higher prevalence in the general population.<sup>7,18,19</sup> As many countries are using the ASQ-3 as the standard developmental screening tool for healthy population screening, it is important to establish its validity and reliability for both detecting developmental and cognitive delay in addition to predicting outcomes at school age.

The BSID-III is equally not without its own confounding factors. Ten percent of the cohort used for establishing normative ranges for each of its three domains; motor, verbal and non-verbal; was comprised of high-risk infants including children with trisomy 21, cerebral palsy, pervasive developmental disorder, premature birth, specific language impairment, prenatal alcohol exposure, birth asphyxiation, small for gestational age.<sup>22</sup> By including high risk infants in the cohort developing the BSID-III, the scores in each domain is right skewed which lowers cut-off values for delay. This increases the likelihood of children with mild to moderate delay remaining undetected by this tool. The ability of the BSID-III to detect intellectual disability has been questioned.<sup>23</sup> This study aimed to assess the domain specific, concurrent validity of the ASQ-3 to predict cognitive delay as detected by the BSID-III at 24 months. We also aimed to assess the predictive ability of both the ASQ-3 and the BSID-III to predict cognitive outcome at school age.

## METHODS

This study was a secondary data analysis of a nested cohort of the Cork BASELINE Birth Cohort Study.<sup>24</sup> Participants were recruited from the Cork BASELINE Birth Cohort Study born between March 2009 and September 2011. BASELINE (<http://www.baselinestudy.net>) was established in 2008 as a follow-up to the Screening for Pregnancy Endpoints (SCOPE) pregnancy study. The study participants, aims and methods of the Cork BASELINE birth cohort study have been previously reported.<sup>25–28</sup> Ethical approval for the Cork BASELINE Birth Cohort Study was granted by the Clinical Research Ethics Committee of the Cork Teaching Hospitals (Ref: ECM 5 (9) 01/ 07/2008), and the study is registered (Ref: NCT01498965) with the United States National Institutes of Health Clinical Trials Registry (<http://www.clinicaltrials.gov>).

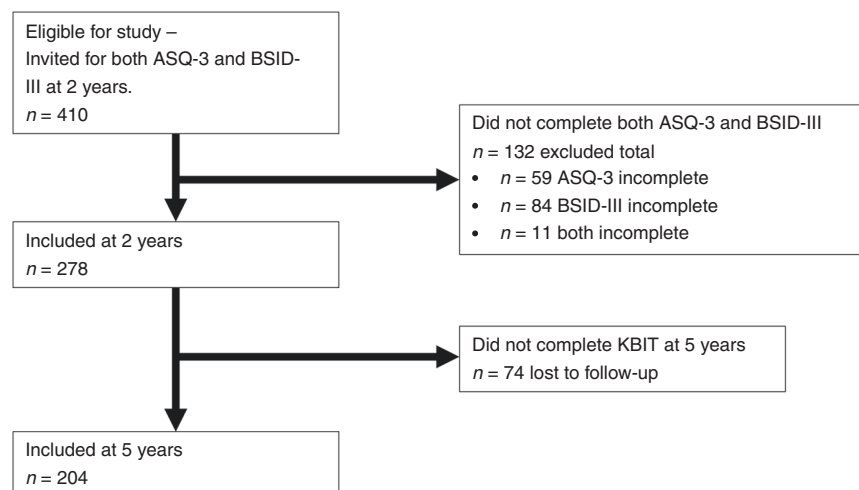
Within the birth cohort, a nested cohort of children who were small or thin for gestational age<sup>28</sup> (SGA and TGA) were invited for additional assessments using the BSID-III at 24 months. TGA refers to infants with a body fat mass <10th percentile. The outcome of these children and their sex and age matched controls has been previously described.<sup>28</sup> Controls were born at term and had birth weights which were appropriate for gestational age (AGA). All BASELINE cohort participants were invited to attend for an IQ assessment at 5 years using the KBIT.

Three developmental assessment tools were utilised in this study; the ASQ-3, BSID-III and KBIT. The Ages and Stages Questionnaire 3 (ASQ-3) is a parentally completed screening tool which can be performed in 10–15 minutes assessing the domains of communication, gross motor, fine motor, problem solving, and personal-social at a variety of ages and in multiple languages.<sup>5</sup> It aims to identify children not achieving age-appropriate developmental milestones across these domains and subsequently refer them for further assessment. Its simplicity and short completion time allow for quick evaluation and identification of children requiring further developmental evaluation.<sup>8</sup> The BSID-III is undertaken by a trained healthcare professional over one to two sessions requiring direct observation of skills in the domains of cognition, language, social-emotional, motor and adaptive behaviour. It is used to identify children with developmental delay who may require intervention services. The KBIT is a professionally administered intelligence quotient assessment which measures verbal and non-verbal intelligence from the ages of 4–90 years.

The BSID-III assessments were administered by research psychologists trained in BSID-III. ASQ-3 questionnaires were posted to the parents the week prior to assessment and they were asked to complete in advance. Any queries were answered on the day of assessment. KBIT was performed by a research nurse trained in the administration of the test. Parents and caregivers were contacted by telephone.

Statistical analysis: The ASQ-3 was compared with the BSID-III across three domains; non-verbal, verbal and motor. The ASQ-3 and the BSID-III were also compared against the KBIT the using verbal and non-verbal domains. The cut-offs which normally trigger further investigation lie one standard deviation (SD) below the population mean (i.e. a score of less than 85) in the BSID-III and KBIT. These cut-off values were adjusted to one SD below the study cohort's mean scores in each domain to ensure a geographically relevant cut-off and to account for the Flynn effect.<sup>29</sup>

For the purposes of our analyses, ASQ-3 scores were treated as pass or fail on each domain. An overall domain score landing in the black area which means that the child needs further assessment was deemed a referral trigger. These "fails" were compared with BSID-III which fell <1 SD below the cohort mean. A score in the grey zone which indicates that a child should be rescreened within six months was not considered a fail for the purposes of this study. When comparing the ASQ-3 and the BSID-III, the fine and gross motor domains for the ASQ-3 were combined to allow direct comparison to the BSID-III motor composite score. If a child triggered a referral for either fine motor, gross motor or both, then this was considered a referral for the purposes of comparison with the BSID-III motor composite score. The communication domain of the ASQ-3 was compared with the language domain of the BSID-III. Finally, the problem-solving score of the



**Fig. 1 Study recruitment flow chart.** Flow chart indicating cohort recruitment and assessment process including reasons for exclusion.

ASQ-3 was compared with the cognitive domain of the BSID-III. The problem-solving score (ASQ-3) and the cognitive domain (BSID-III) represent non-verbal for the purposes of our analyses and were compared with non-verbal scores for the KBIT.

The ASQ-3, the BSID-III and the KBIT were analysed for their levels of agreement and suitability as screening and assessment tools. Statistical analyses were completed using IBM SPSS Version 26 software. Descriptive statistics, Pearson correlations and concurrent validity and predictive ability were measured using sensitivity, specificity and area under receiver operating characteristic curve (AUROC). Sensitivity and specificity were used instead of positive and negative predictive value as this nested cohort of children who were small or thin for gestational age is at higher risk of developmental delay than the general population.

## RESULTS

Four hundred and ten subjects were invited for both ASQ-3 and the BSID-III at 24 months. Participant flow chart and reasons for exclusion are outlined in Fig. 1. The mean (SD) birthweight of the children studied was 3445 (506) g, with a male-to-female ratio was 52:48. The mean(SD, range) for BW was 3445.58 (506.619, 3970), gestational age 39.56 (1.489, 13) weeks. Mean (SD) age assessment for ASQ and BSID-III respectively was 25.4 (1.4) and 26.9 (1.7) months. 13.5% (37/274, four values missing) were SGA. ASQ-3 and BSID-III assessments took place on the same day in 83.8% of cases, with some assessments taking 2–3 sessions to complete.

The subjects excluded due to non-completion of both assessments were compared with those who completed both (ASQ-3 and BSID-III). No significant differences were observed between the two groups (completed versus non-complete data groups) in their gestational age in weeks; sex distribution; maternal education; maternal marital status and maternal occupation. Figure 1 illustrates the attrition rate and numbers included at each stage of the study.

Outcome at 24 months: At 24 months, the mean(SD) BSID-III cognitive composite score of the cohort was 98.63 (11.30). The cohort specific cut-off of <1 SD was calculated as 87, with 29/278 (10.4%) children having a score below this cut-off. At 24 months, only 14/278(5%) children failed the problem-solving domain of the ASQ-3 including 8 false positives. For explanation of each domain's geographically relevant cut-off (please see Table 1).

**Table 1.** Geographically relevant cut-off's for BSID-III and KBIT.

Geographically relevant cut-off's for BSID-III and KBIT			
	Mean	Standard Deviation (SD)	Mean < 1 SD
BSID-III Cognitive	98.63	11.30	87.33 (87)
BSID-III Language	106.65	15.29	91.36 (91)
BSID-III Motor	104.54	11.88	92.66 (92)
KBIT Verbal	107.34	9.54	97.8 (98)
KBIT Non-verbal	100.80	9.73	91.07 (91)

Cut-off scores based on 1SD below the cohort mean performance in each domain.

**Table 2.** Concurrent validity of the ASQ-3 domain-specific clinical cut-offs and the language, motor and cognitive domains respectively of the Bayley Scales of Infant Development Version III (BSID-III).

	N	True Neg Pass both	False neg Pass ASQ-3, <1SD BSID	False Pos Clinical range ASQ-3/ Pass BSID	True Pos Clinical range ASQ-3/<1SD BSID
Cognitive	278	241	23	8	6
Communication	278	233	33	1	11
Motor	278	181	20	57	20

The concurrent validity data of the ASQ-3 and BSID-III at 24 months are displayed on Table 2. Of the 29/278(10.4%) children who had a BSID-III cognitive score <1 SD at 24 months, only 6/29 also had a fail score on the problem-solving domain of the ASQ-3 giving a sensitivity for the detection of cognitive delay of 20.7%. The ability of the ASQ-3 domains to detect language or motor delay was higher, with sensitivities of 25% and 50%, respectively. Specificities for cognitive, communication and motor domains were 96.8%, 99.5% and 76.1%, respectively. Overall predictive ability is displayed in Table 3. Again, whilst the motor and language domains of the ASQ-3 showed a reasonable predictive ability for a motor score <1 SD on the BSID-III, the predictive ability of the problem-solving domain for the cognitive scale of the BSID-III was poor.

The overall predictive ability of the ASQ-3 differed with severity of delay (Table 4). For the 55 children with a BSID-III score <1 SD in any domain, 25 (45.5%) also failed at least one element of the ASQ. For the 13 with more severe delay (<2 SD) on any domain of the BSID-III, 11(84.6%) also failed at least one element of the ASQ.

Lastly, we examined the ability of both the ASQ and BSID-III at 24 months to predict cognitive difficulties at 5 years. Defining low average IQ as <1 SD for either verbal or non-verbal IQ on the KBIQ, 36/204(17.6%) had a low average verbal KBIT score and 24/204(11.8%) had a low average non-verbal KBIT score. The KBIT IQ composite score n, mean, (SD) = 204, 105.14, (8.88). The correlation between the BSID-III domains and KBIT scores at 5 years is displayed in Table 5. The cognitive composite score of the BSID-III correlated with both the verbal and non-verbal scores, and the Total IQ composite score at 5 years. The best correlation was seen between the language composite score at two years and the verbal score at 5 years. The ability of a low average ASQ and BSID-III to predict a low average IQ at 5 years is depicted in Tables 6 and 7. In all cases, high specificity was offset by low sensitivity. The best predictor of overall IQ at 5 years was the composite language score of the BSID-III at 2 years (Table 6).

## DISCUSSION

We have shown that the ASQ-3 has poor sensitivity for detecting mild-moderate delay when compared with the BSID-III at 24 months. Both the ASQ-3 and BSID-III demonstrate poor ability to predict low average cognitive outcomes at school age. However, BSID-III has slightly better sensitivities and significant AUROC's whereas ASQ-3's AUROCs are not statistically significant. The motor and language domains of the ASQ-3 performed best, whilst the cognitive domain showed the lowest concurrent validity and predictive ability at both time-points. The specificity of the ASQ-3 was high in all domains. This confirms the ability of the ASQ-3 to identify normative development, but also highlights that many children with cognitive difficulties at school age will be missed in countries relying on the ASQ-3 as a screening tool. Similar profiles of low sensitivity and high specificity have been reported for the ASQ-2 when compared with the Bayley Scales.<sup>19</sup>

The ASQ is one of the most widely used screening assessments of early child development, forming part of the recommended screening schedule in many countries, including the public health assessment protocol in Ireland, where this study was developed.<sup>30</sup> The ease of administration, requiring care-giver report only, makes it

**Table 3.** Comparison of domain-specific sensitivities and specificities for ASQ-3 to detect an abnormal BSID-III (<1SD cohort mean).

ASQ-3	N	BSID-III relevant domain < 1SD cohort mean				p value
		Sensitivity	Specificity	AUROC	Range	
Cognitive	278	20.7%	96.8%	0.587	(0.467–0.708)	0.124
Communication	278	25%	99.5%	0.623	(0.521–0.724)	0.010
Motor	278	50%	76.1%	0.630	(0.532–0.728)	0.008

**Table 4.** Ability of ASQ-3 to detect an abnormal BSID-III at 24 months with stratification for type of delay based on all-domain performance (Cognitive, communication and motor).

ASQ-3	BSID-III (all domain performance)			Total
	Severe (<2SD)	Mild-Moderate (<1SD)	Expected	
Fail Any Domain	11 (84.6%)	25 (45.5%)	46 (21.9%)	<b>80</b>
Pass All Domains	2 (15.4%)	30 (54.5%)	164 (78.1%)	<b>198</b>
Total	<b>13</b>	<b>55</b>	<b>210</b>	<b>278</b>

Mild-moderate delay is characterised as between 1SD and 2SD below the cohort mean. Severe delay is characterised by <2SD below the cohort mean.

**Table 5.** Correlation between Bayley Scales of Infant development-Version III (BSID-III) assessment at 2 years and Kaufmann Brief IQ test at 5 years.

BSID-III vs KBIT		Total verbal standard score	Total non-verbal standard score	IQ composite score
Language composite score	R (p value)	0.435 P < 0.001	0.158 P = 0.024	0.365 P < 0.001
Motor composite score	R (p value)	0.206 P = 0.003	0.110 P = 0.117	0.207 P = 0.003
Cognitive composite score	R (p value)	0.388 P < 0.001	0.342 P < 0.001	0.463 P < 0.001

R = Pearson's correlation coefficient.

attractive for widespread screening, particularly in low-resource countries. Adequate reliability and validity of the ASQ have been demonstrated in U.S. populations and other high-resource countries.<sup>31</sup> Validity reported by the developers of the ASQ has used comparisons to early intervention/early childhood special education eligibility evaluations using the Battelle Developmental Inventory in high-risk groups referred for further assessment.<sup>14</sup> These groups consisted of more than 50% deemed eligible for special education, making it difficult to compare this to the widespread use of the ASQ as a developmental screener in a healthy population. Data on the predictive ability of the ASQ has been reported in a number of countries, using locally adjusted versions of the ASQ in non-English speaking cohorts with calculated ROC ranging from 0.66 to 0.87.<sup>32</sup> Data in English-speaking populations are limited with a systematic review having identified 32 publications related to this from a combined ten cohorts; of these ten cohorts only one used the English language ASQ.<sup>33</sup>

Although the ASQ is designed to give a global impression of the child's development, the assessment is divided into individual domains. Our focus for this study was the prediction of cognitive delay. Whilst language and motor assessment concurrent validity between the ASQ and the BSID-III was more robust, detection of mild-moderate cognitive delay was very poor. We have also shown that the ASQ-3 is more effective at detecting severe delay than mild or moderate delay, in keeping with previous reports from Sheldrick et al.<sup>34</sup> and Gollenberg et al.<sup>18</sup>. When one looks at all-domain performance, 84.6% of those classified by the BSID-III as having severe delay were detected by the ASQ-3. Only 45.5% were detected in the mild to moderate category. Often, children with severe delay will likely be detected on routine observation by parents and medical professionals without the requirement for validated screening

tools. Most children with cognitive delay fall into the mild to moderate<sup>34</sup> category which is the cohort least likely to be detected by the ASQ-3 alone. These children are also the group who may benefit most from early detection and implementation of early intervention services.

This study focused on assessing the ASQ-3's performance at two years, as recommended in Ireland's national public health screening schedule, and prediction of cognitive outcomes at school-age. The communication and problem-solving domains of the ASQ-3 were assessed for their ability to predict future performance in the verbal, non-verbal and total intelligence domains of the KBIT respectively. The incidence of delay was similar to that seen in the total BASELINE cohort at 5 years.<sup>27,35</sup> It should be noted that although KBIT is an abbreviated IQ assessment and is not a comprehensive test. It has demonstrated good reliability in assessing cognitive outcomes at school age compared to formal testing using either WISC-IV General Ability Index or the WAIS-III. It has been used in a number of large birth cohorts due to its ease of administration. It does not however give a detailed cognitive profile allowing separation of individual cognitive functions. Nor does it predict specific learning disabilities.<sup>36–38</sup> Whilst the ASQ-3 showed excellent specificity, the sensitivity for the detection of low average non-verbal IQ was very low. This raises questions regarding the utility of the ASQ-3 as a screening tool for cognitive delay. We require screening tools to identify children suitable for further testing. As it is currently scored, the ASQ-3 has such high specificity that further testing is almost not required, whilst its low sensitivity means that the majority of cases of cognitive delay will be missed.

The BSID-III performed better than the ASQ-3 for prediction of outcome at 5 years. This is to be expected with a directly administered test. The overall prediction of both verbal and non-verbal IQ was acceptable, with AUROCs between 0.695 and 0.723.

**Table 6.** Ability of ASQ-3 and BSID-III to predict <1SD below cohort mean performance in KBIT at 5 years of age in both verbal and non-verbal domains.

	AUROC (95% CI)	p value	Sensitivity	Specificity
Prediction of verbal IQ < 1SD in KBIT				
ASQ-3 communication	0.536(0.428–0.644)	0.502	8.3%	98.8%
BSID-III language	0.723(0.63–0.81)	<0.001	22.2%	88.7%
Prediction of non-verbal IQ < 1SD in KBIT				
ASQ-3 problem solving	0.528 (0.400–0.655)	0.659	8.3%	97.2%
BSID-III cognitive	0.695(0.59–0.80)	0.002	20.8%	94.4%
Prediction of overall IQ < 1SD in KBIT				
BSID-III language	0.732(0.64–0.82)	<0.001	15.6%	94.2%
BSID-III cognitive	0.696(0.61–0.79)	<0.001	18.8%	94.8%

**Table 7.** Domain-specific ability of ASQ-3 and BSID-III to predict <1SD below cohort mean performance in KBIT composite score at 5 years of age.

Prediction of KBIT composite score < 1SD		
	Sensitivity	Specificity
ASQ-3 problem-solving	12.5%	98.2%
BSID-III cognitive	15.6%	94.2%
ASQ-3 communication	9.4%	98.8%
BSID-III language	21.9%	88.4%
ASQ-3 motor	25%	75%
BSID-III motor	28.1%	88.4%

However, this was driven by very high specificity, meaning that the majority of children with low average IQ scores at 5 years were not detected by a BSID-III at 24 months. It is estimated that less than half of children with developmental delay are detected prior to school entrance, with the vast majority of those detected receiving no intervention in the very early years.<sup>39,40</sup> Our study supports the need for better early assessments of cognitive ability and executive function to allow useful intervention and support prior to school entry. Both assessments were performed once at 24 months and it may be that repeated measures to assess the child's trajectory over time may improve prediction.<sup>41</sup>

It is not surprising that the ASQ assessments based on caregiver report perform better in the quantification of tangible skills such as speech and motor milestones. As with all parentally reported assessments, there is an inherent bias, which may be influenced by the parents own socio-economic background, parenting experience or education level. Care-givers assessment of a child's early problem-solving skills is inherently more difficult. However, it is these early executive functions which are most highly predictive of the lifetime course, opportunities and health of that child.<sup>1</sup> We must recognise that current screening, whether by caregiver report or direct developmental assessment does not detect the majority of cases of low average cognitive ability. New methods of early assessment are required if we are to give these children opportunities to reach their full potential. This particularly applies to mild-moderate delay where early intervention may improve that child's academic attainment and consequently their future quality of life.<sup>42–46</sup>

We have focused on examining the agreement between neurodevelopmental performance at 24 months and subsequent performance at school age as data on predictive performance is scarce.<sup>33</sup> This study utilised BSID-III as the gold standard at two years of age. It should be noted that BSID-IV was released in 2019. It should be noted that the Bayley Scales of Infant and Toddler Development IV did not include the high-risk populations in the normative group as was the case for the BSID-III. Outcomes at five

years of age will not be available with the BSID-IV for several years. The BSID-IV has updated its administrative protocol for ease of use and to shorten the assessment, without altering expected scores. Thus we feel that our findings are relevant for current use of either BSID-III or BSID-IV.<sup>47</sup> There were some incomplete data in this study as patients were excluded due to non-completion of either the ASQ-3, the BSID-III or both. Subsequently, 74 further patients did not complete the school-age assessment at five years. However, the children with incomplete data did not differ from the group in their demographic or socioeconomic variables. The amount of time taken to complete the BSID-III may be a barrier to its completion with 45/278 requiring more than one session to complete the assessment. Each session may take 60–90 min. However, each subject was afforded the opportunity of multiple sessions to complete the BSID-III assessment. Our study was not focused on examining the acceptability or ease of use of both methods of assessment, but instead a direct comparison of detection of domain-specific delay.

## CONCLUSION

In summary, the parentally completed ASQ-3 has poor concurrent validity with the cognitive scales of the BSID-III at the same age. It is more effective at identifying children with severe developmental delay but does not achieve sensitivities requisite with an effective screening tool. Both the ASQ-3 and the BSID-III demonstrate poor ability to predict cognitive ability at school-age in both verbal and non-verbal domains. Both tests have high specificity and are adept at predicting a normal performance at 5 years. The majority of children with low average IQ at school age will not be detected using current screening methods.

## DATA AVAILABILITY

De-identified individual participant data (including data dictionaries) will be made available, in addition to study protocols, the statistical analysis plan, and the informed consent form. The data will be made available upon publication to researchers who provide a methodologically sound proposal for use in achieving the goals of the approved proposal. Proposals should be submitted to d.murray@ucc.ie

## REFERENCES

- Caspi, A. et al. Childhood forecasting of a small segment of the population with large economic burden. *Nat. Hum. Behav.* **1**, 0005, <https://doi.org/10.1038/s41562-016-0005> (2016).
- Ramey, C. T. & Ramey, S. L. Prevention of intellectual disabilities: early interventions to improve cognitive development. *Prev. Med.* **27**, 224–232 (1998).
- Spittle, A., Orton, J., Anderson, P. J., Boyd, R. & Doyle, L. W. Early developmental intervention programmes provided post hospital discharge to prevent motor and cognitive impairment in preterm infants. *Cochrane Database Syst. Rev.* **2015**, Cd005495 (2015).
- ICD-11. Neurodevelopmental disorders. World Health Organisation. 2020 accessed 25 Aug 2020. <https://icd.who.int/dev11/f/en-/http%3a%2f%2fid.who.int%2f%2fid%2f%2f1516623224>

5. Ages and Stages. An Introduction to ASQ-3. Paul H Brookes Publishing Co; 2014. <https://agesandstages.com/wp-content/uploads/2015/02/Training-PPT-for-website-10-3-14.pdf>
6. Glascoe, F. P. Early detection of developmental and behavioral problems. *Pediatr. Rev.* **21**, 272–279 (2000). quiz 280.
7. Health Service Executive. Training Programme for public health nurses and doctors in child health screening, surveillance and health promotion: developmental assessment. pp. 4–34 (Health Service Executive, 2005).
8. Vitrikas, K., Savard, D. & Bucaj, M. Developmental delay: when and how to screen. *Am. Fam. Physician* **96**, 36–43 (2017).
9. De Jesus, L. C. et al. Outcomes of small for gestational age infants born at <27 weeks' gestation. *J. Pediatr.* **163**, 55–60.e3 (2013).
10. Pierrat, VM-M. et al. The EPIPAGE-2 Writing Group. Neurodevelopmental outcome at 2 years for preterm children born at 22 to 34 weeks' gestation in France in 2011: EPIPAGE-2 cohort study. *BMJ*. **358**, 1–13 (2017). <https://doi.org/10.1136/bmj.j3448>.
11. Synnes, A. et al. Determinants of developmental outcomes in a very preterm Canadian cohort. *Arch. Dis. Child. - Fetal Neonatal Ed.* **102**, F235 (2017).
12. Council on Children With Disabilities; Section on Developmental Behavioral Pediatrics; Bright Futures Steering Committee; Medical Home Initiatives for Children With Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. *Pediatrics*. **118**, 405 (2006).
13. Veldhuizen, S., Clinton, J., Rodriguez, C., Wade, T. J. & Cairney, J. Concurrent validity of the Ages And Stages Questionnaires and Bayley Developmental Scales in a general population sample. *Acad. Pediatr.* **15**, 231–237 (2015).
14. Squires, J., Bricker, D., Twombly, E., Potter, L. ASQ Technical Report. 2009. 2021. Accessed 17 Oct 2021. [https://agesandstages.com/wp-content/uploads/2017/05/ASQ-3-Technical-Appendix\\_web.pdf](https://agesandstages.com/wp-content/uploads/2017/05/ASQ-3-Technical-Appendix_web.pdf)
15. Bian, X. et al. Translation and use of parent-completed developmental screening test in Shanghai. *J. Early Child. Res.* **10**, 162–175 (2012).
16. Velikonja, T. et al. The psychometric properties of the Ages & Stages Questionnaires for ages 2–2.5: a systematic review. *Child Care Health Dev.* **43**, 1–17 (2017).
17. Woodward, B. J. et al. Use of the Ages and Stages Questionnaire and Bayley Scales of Infant Development-II in neurodevelopmental follow-up of extremely low birth weight infants. *J. Perinatol.* **31**, 641–646 (2011).
18. Gollenberg, A. L., Lynch, C. D., Jackson, L. W., McGuinness, B. M. & Msall, M. E. Concurrent validity of the parent-completed Ages and Stages Questionnaires, 2nd Ed. with the Bayley Scales of Infant Development II in a low-risk sample. *Child Care Health Dev.* **36**, 485–490 (2010).
19. Lindsay, N. M., Healy, G. N., Colditz, P. B. & Lingwood, B. E. Use of the Ages and Stages Questionnaire to predict outcome after hypoxic-ischaemic encephalopathy in the neonate. *J. Paediatr. Child Health* **44**, 590–595 (2008).
20. Simard, M. N., Luu, T. M. & Gosselin, J. Concurrent validity of ages and stages questionnaires in preterm infants. *Pediatrics* **130**, e108–e114 (2012).
21. Skellern, C. Y., Rogers, Y. & O'Callaghan, M. J. A parent-completed developmental questionnaire: follow up of ex-premature infants. *J. Paediatr. Child Health* **37**, 125–129 (2001).
22. Oakland T. A., Glen P. A., Weiss L. G.. Bayley-III Clinical Use and Interpretation. Elsevier; 2010.
23. Moore, T., Johnson, S., Haider, S., Hennessy, E. & Marlow, N. Relationship between test scores using the second and third editions of the Bayley Scales in extremely preterm children. *J. Pediatr.* **160**, 553–558 (2012).
24. O'Donovan, S. M. et al. Cohort profile: The Cork BASELINE Birth Cohort Study: babies after SCOPE: evaluating the longitudinal impact on neurological and nutritional endpoints. *Int. J. Epidemiol.* **44**, 764–775 (2015).
25. McCarthy, E. K. et al. Behavioral consequences at 5 y of neonatal iron deficiency in a low-risk maternal-infant cohort. *Am. J. Clin. Nutr.* **113**, 1032–1041 (2021).
26. McCarthy, E. K. et al. Antenatal vitamin D status is not associated with standard neurodevelopmental assessments at age 5 years in a well-characterized prospective maternal-infant cohort. *J. Nutr.* **148**, 1580–1586 (2018).
27. Lenehan, S. M. et al. The impact of short-term predominate breastfeeding on cognitive outcome at 5 years. *Acta Paediatr.* **109**, 982–988 (2020).
28. O'Neill, S. M. et al. Thin-for-gestational age infants are at increased risk of neurodevelopmental delay at 2 years. *Arch. Dis. Child.-Fetal Neonatal Ed.* **102**, F197–F202 (2017).
29. Williams, R. L. Overview of the Flynn effect. *Intelligence* **41**, 753–764 (2013).
30. Squires J., Bricker D., Twombly E., Potter L. *Ages & Stages Questionnaires: A Parent-Completed, Child-Monitoring System (3rd ed.)*. (Paul H. Brookes Publishing; 2009).
31. Agarwal, P. K. et al. Prospective evaluation of the Ages and Stages Questionnaire 3rd Edition in very-low-birthweight infants. *Dev. Med. Child. Neurol.* **59**, 484–489 (2017).
32. Marks, K. et al. The thorny nature of predictive validity studies on screening tests for developmental-behavioral problems. *Pediatrics* **122**, 866–868 (2008).
33. Schonhaut, L., Maturana, A., Cepeda, O. & Serón, P. Predictive validity of developmental screening questionnaires for identifying children with later cognitive or educational difficulties: a systematic review. *Front. Pediatr.* **9**, 698549 (2021).
34. Sheldrick, R. C., Marakovitz, S., Garfinkel, D., Carter, A. S. & Perrin, E. C. Comparative accuracy of developmental screening questionnaires. *JAMA Pediatr.* **174**, 366–374 (2020).
35. Bowe, A. K., Hourihane, J., Staines, A. & Murray, D. M. The predictive value of the ages and stages questionnaire in late infancy for low average cognitive ability at age 5. *Acta Paediatr.* **111**, 1194–1200 (2022).
36. Bain, S. K. et al. Kaufman Brief Intelligence Test, Bloomington, MN: Pearson, Inc. *J. Psychoeduc. Assess.* **28**, 167–174 (2010). (2004).
37. Oken, E. et al. Cohort profile: project viva. *Int. J. Epidemiol.* **44**, 37–48 (2015).
38. Mahmassani, H. A. et al. Maternal diet quality during pregnancy and child cognition and behavior in a US cohort. *Am. J. Clin. Nutr.* **115**, 128–141 (2022).
39. Morelli, D. L. et al. Challenges to implementation of developmental screening in urban primary care: a mixed methods study. *BMC Pediatr.* **14**, 1–11 (2014).
40. Dearlove, J. & Kearney, D. How good is general practice developmental screening? *Br. Med. J.* **300**, 1177–1180 (1990).
41. O'Connor, C. M., Ryan, C. A., Boylan, G. B. & Murray, D. M. The ability of early serial developmental assessment to predict outcome at 5 years following neonatal hypoxic-ischaemic encephalopathy. *Early Hum. Dev.* **110**, 1–8 (2017).
42. Sharkey, M. A. et al. The effect of early referral and intervention on the developmentally disabled infant: evaluation at 18 months of age. *J. Am. Board Fam. Pract.* **3**, 163–170 (1990).
43. Gombay, D. S., Larner, M. B., Stevenson, C. S., Lewit, E. M. & Behrman, R. E. Long-term outcomes of early childhood programs: analysis and recommendations. *Future Child.* **5**, 6–24 (1995).
44. Shonkoff, J. P. & Hauser-Cram, P. Early intervention for disabled infants and their families: a quantitative analysis. *Pediatrics* **80**, 650–658 (1987).
45. Ramey, C. T. & Ramey, S. L. Which children benefit the most from early intervention? *Pediatrics* **94**, 1064 (1994).
46. Scherzer, A. L., Mike, V. & Ilson, J. Physical therapy as a determinant of change in the cerebral palsied infant. *Pediatrics* **58**, 47–52 (1976).
47. Bayley N., Aylward, G. Bayley Scales of Infant and Toddler Development. Fourth edition (Bayley-4). (Bloomington, MN: NCS Pearson, 2019).

## AUTHOR CONTRIBUTIONS

C.D. and D.M. conceptualised and designed the study, drafted the initial manuscript and reviewed and revised the manuscript. They also carried out the initial analyses. A.I., J.O.B.H. and M.E.K. planned and oversaw all aspects of the BASELINE Birth Cohort study, oversaw data collection on the original cohort and critically reviewed the manuscript. All authors approved the final manuscript and agree to be accountable for all aspects of the work.

## FUNDING

The Cork BASELINE Birth Cohort Study was funded by the National Children's Research Centre through project grants January 2009 and October 2012. Open Access funding provided by the IReL Consortium.

## COMPETING INTERESTS

The authors have no conflicts of interest relevant to this article to disclose.

## CONSENT STATEMENT

Parental or guardian consent was received for each subject involved in this study. The data will be made available upon publication to researchers who provide a methodologically sound proposal for use in achieving the goals of the approved proposal. Proposals should be submitted to d.murray@ucc.ie.

## ADDITIONAL INFORMATION

**Correspondence** and requests for materials should be addressed to Cian Duggan.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in

the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023