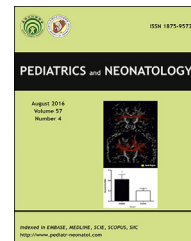


Available online at www.sciencedirect.com

ScienceDirect

journal homepage: <http://www.pediatr-neonatol.com>

Original Article

Concurrent validity of the ages and stages questionnaires with Bayley Scales of Infant Development-III at 2 years – Singapore cohort study

Pratibha Keshav Agarwal ^{a,*}, Huichao Xie ^b,
Anu Sathyan Sathyapalan Rema ^c, Michael J. Meaney ^d,
Keith M. Godfrey ^e, Victor Samuel Rajadurai ^f,
Lourdes Mary Daniel ^a

^a Dept of Child Development; KK Women's and Children's Hospital Singapore, Singapore

^b Psychology and Child & Human Development, Nanyang Technological University National Institute of Education, Singapore

^c Agency for Science Technology and Research (A*STAR), Singapore

^d McGill University, Canada

^e MRC Life Course Epidemiology Centre and NIHR Southampton Biomedical Research Centre, University of Southampton and University Hospital Southampton NHS Foundation Trust, UK

^f Dept of Neonatology, KK Women's and Children's Hospital Singapore, Singapore

Received Aug 2, 2022; received in revised form Feb 9, 2023; accepted Mar 2, 2023

Available online ■ ■ ■

Key Words

ASQ-3;
Bayley-III;
cohort study;
developmental
screening;
validity

Background: With increasing acceptance of universal developmental screening in primary care, it is essential to evaluate the local validity and psychometric properties of commonly used questionnaires like the parent-completed Ages and Stages Questionnaires, 3rd Edition (ASQ-3) in identifying developmental delays. The aim of this study is to assess the convergent validity of the ASQ-3 with the Bayley Scales of Infant Development-3rd edition (Bayley-III) in identifying developmental delay in a low-risk term cohort in Singapore.

Methods: ASQ-3 and Bayley-III data was collected prospectively with generation of ASQ-3 cut-off scores using three different criteria: 1-standard deviation (SD) (Criterion-I) or 2-SD (Criterion-II) below the mean, and using a Receiver Operator Curve (ROC) (Criterion-III). Sensitivity, specificity, positive (PPV) and negative (NPV) predictive values were calculated. Correlations between the ASQ-3 and Bayley-III domains were evaluated using Pearson coefficients.

Results: With all three criteria across different domains ASQ-3 showed high specificity (72–99%) and NPV (69–98%), but lower sensitivity (19–74%) and PPV (11–59%). Criterion-I

* Corresponding author.

E-mail address: pratibha.agarwal@singhealth.com.sg (P.K. Agarwal).

<https://doi.org/10.1016/j.pedneo.2023.03.013>

1875-9572/ Copyright © 2023, Taiwan Pediatric Association. Published by Elsevier Taiwan LLC. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Please cite this article as: P.K. Agarwal, H. Xie, A.S. Sathyapalan Rema et al., Concurrent validity of the ages and stages questionnaires with Bayley Scales of Infant Development-III at 2 years – Singapore cohort study, Pediatrics and Neonatology, <https://doi.org/10.1016/j.pedneo.2023.03.013>

identified 11–21% of children as “at-risk of developmental delay,” and was the most promising criterion measure, with high specificity (82–91%), NPV (69–74%) and overall agreement of 64–71%. Moderate-strong correlations were seen between ASQ-3 Communication and Bayley-III Language scales ($r = 0.44–0.59$, $p < 0.01$). The lowest sensitivities were seen in the motor domains.

Conclusions: ASQ-3 is reliable in low-risk settings in identifying typically developing children not at risk of developmental delay, but it has modest sensitivity. Moderate-strong correlations seen in the communication domain are clinically important for early identification of language delay, which is one of the most prevalent areas of early childhood developmental delay.

Copyright © 2023, Taiwan Pediatric Association. Published by Elsevier Taiwan LLC. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The first 1000 days of life are recognized as being vital in a child’s development as a consequence of the plasticity and vulnerability of the developing brain, with a strong influence of risk and protective factors on subsequent outcomes into school age and adult life. Early identification of and intervention for developmental delays in this period are critical in optimizing outcomes, both in childhood and into adult life. Deleterious consequences in education, health, earning potential, and socio-emotional domains have been reported for those with delayed intervention.^{1–3}

However, identification is often delayed, with no more than 30% of children with developmental problems being identified before school age.⁴ Given that the prevalence of developmental delay in preschool children is 5–15%,^{5–7} developmental screening is increasingly recommended in primary care public health systems. The American Academy of Pediatrics (AAP)⁸ recommends developmental surveillance at all pediatric visits and formal screening at ages 9, 18, and 30 months. Evidence-based standardized screening tools have been shown to increase the detection of possible developmental delays from 16% to 62%.⁹ Thus, having a valid, easy-to-use tool to assess development in the early years is critical.

Traditionally, standardized instruments such as the Bayley Scale of Infant and Toddler Development, Third Edition (Bayley-III) have been widely considered the “gold standard” in diagnosing developmental delay in children between 0 and 42 months. However, the Bayley-III requires individual administration by trained professionals with constraints of time, cost, and resources. Efforts have thus been made towards effective use of screening approaches in low-risk populations for accurate identification of children at risk of delay as well as reduction of unnecessary referrals and expensive diagnostic assessments in children with normal screening results. In recent years, the use of parent-completed screeners, such as the Parents’ Evaluation of Developmental Status¹⁰ and the Ages and Stages Questionnaire—Third Edition,¹¹ have become increasingly popular due to their psychometric properties, lower cost, and ease of use with minimal training.

Parent-led developmental screening may also enhance communication between parents and healthcare providers and encourage parents to be active partners in assessing

their child’s needs and progress with involvement in subsequent evaluations and interventions, which are essential for optimizing outcomes.¹²

Our previous work has shown the ASQ-3 to be a valuable screening tool for the neurodevelopmental follow-up of preterm, very low birth weight infants (VLBW)¹³ and a reliable developmental screening tool in a low-risk cohort in Singapore.¹⁴ However, further evidence for ASQ-3’s performance in multi-ethnic populations is still needed. The aim of the present study was to evaluate the concurrent validity of ASQ-3 screening with the Bayley-III in a low-risk, multi-ethnic cohort in Singapore.

2. Methods

The current study was part of a parent-offspring longitudinal research cohort, the Growing Up in Singapore Towards Healthy Outcomes (GUSTO) project.¹⁵ GUSTO recruited 1247 pregnant women in their first trimester and collected outcome data on development, growth, and other parameters throughout the antenatal, neonatal and early childhood periods for 1176 children born between November 2009 and May 2011.

The current sub-study recruited participants from the neurodevelopmental domain arm of the larger GUSTO cohort. ASQ-3 forms were sent to parents or caregivers to complete and return at the subsequent clinic visit. Trained examiners administered the Bayley-III to toddlers at age 24 months. The study was approved by Institutional Review Boards (CIRB Ref: 2009/280/D) (DSRB Ref: 09/021), and consent was obtained from all parents.

2.1. Measures

2.1.1. Ages & stages questionnaires, 3rd edition (ASQ-3)⁽¹¹⁾

The ASQ-3 is a parent-completed developmental screening tool for children aged 1–66 months, with 21 questionnaires targeting specific age ranges. It measures development in five domains: communication (CM), gross motor (GM), fine motor (FM), problem solving (CG), and personal-social (PS). Each questionnaire has six scoring items per developmental domain using a three-level system: “yes” (10 points), “sometimes” (5 points), and “not yet” (0 points). Total domain scores are computed by adding points obtained

from each domain item and comparing them to cut-off scores from a normative sample. A child with an ASQ-3 score below the cut-off point is considered “at risk” for developmental delays and in need of referral for further evaluation. The questionnaires have been translated into Chinese, Malay, and Tamil for non-English speaking parents/caregivers.¹⁴

2.1.2. Bayley-III¹⁶

The Bayley-III is a psychometric assessment test measuring three developmental domains: cognitive, language (i.e., receptive and expressive), and motor (i.e., gross and fine motor) skills in children aged 0.5–42 months. It yields a standardized composite score with a mean of 100 and a standard deviation (SD) of 15. A composite score below 70 indicates significant developmental delay, while a score of 70–84 is interpreted as “suspect delay.”

The Bayley-III was administered at home to 2-year-old children by psychologists/examiners who spoke two or more of the four languages. Examiners attended a three-day training session and completed 10 h or more of supervised administration.

2.2. Data analysis

Three sets of local ASQ-3 cut-off criteria were first generated using two methods, either the mean and less than one (Criterion-I) or two (Criterion-II) standard deviations (SD), or using a receiver operating characteristic (ROC) curve in which the Bayley-III was the criterion measure (Criterion-III). These three cut-off values were then compared for agreement and concurrent validity with the Bayley-III.

The data were analyzed using IBM SPSS version 26. The analysis proceeded in five steps. First, descriptive statistics for both tools (e.g., mean and SD, proportions) of the study sample were calculated. Second, the categorization results (i.e., typical development, developmental delay) were determined for the ASQ-3 and Bayley-III. As indicated in our previous study,¹⁴ and also in a Dutch study,¹⁷ U.S. norms for the ASQ-3 might not be adequate when used in different populations. In the current study, our local sample ($n = 335$) was used to generate cut-off points with 1 and 2 SD values for the ASQ-3 and Bayley-III. The third step required using the ROC analyses to determine optimal cut-off points for sensitivity and specificity using the Bayley-III as a reference measure. A Bayley-III domain composite score of <70 was coded as “developmental delay.” For the ROC analyses, scores below the local cut-off on any of the three Bayley-III scales (i.e., language, cognitive, and motor) were coded as “developmental delay.” An area under the curve (AUC) of >0.60 was considered acceptable¹⁸ for generating cut-off points.

In the fourth step, the validity of the ASQ-3 cut-off points were evaluated based on sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV). Sensitivity and specificity values above 70% were considered acceptable as recommended by AAP Pediatrics⁸ and another study by Steenis et al.¹⁹

Lastly, correlations between ASQ-3 and Bayley-III scores in each domain were analyzed using Pearson correlation coefficients. Correlations >0.50 were considered high,

0.30–0.50 moderate, and below 0.30 low. We expected ASQ-3 and Bayley-III correlations to be moderate or high in convergent domains (e.g., ASQ-3 communication and Bayley-III language), and low (i.e., $r < 0.30$) in divergent domains (e.g., ASQ-3 CG and Bayley-III GM).

3. Results

3.1. Participants

Children ($n = 335$) with complete data were included in this study; their demographic characteristics were compared to national population data²⁰ as shown in Table 1, revealing a significant difference in income distribution, ethnicity, and maternal education. Minority ethnicities were more commonly seen, and fewer higher income families ($>SGD 4000$) were reported in our cohort compared to national data.²⁰ As shown in Table 1, 74.5% of the mothers had education at or higher than college level, compared to 42.40% of the national population.

The mean age of the participants was 24.19 months ($SD = 0.40$, range 23–25 months). Their mean gestational age and birth weight were 38.84 weeks ($SD = 1.34$) and 3101.02 g ($SD = 431.40$), respectively, with 50.5% being boys.

3.2. Descriptive statistics of the ASQ-3 and Bayley-III

Mean scores in the different domains and total ASQ-3 and Bayley-III results at age 24 months are presented in Table 2. Bayley-III standard composite scores in each domain were normed on the local sample with a mean of 100 ($SD 15–16$). Bayley-III scores $\leq 2SD$ below the mean were seen in 9 (2.71%), 12 (3.6%), and 8 (2.4%) children in the language,

Table 1 Demographic characteristics of the study cohort ($n = 335$) and comparison with Singapore’s national population.

Characteristics	n (%) in the study	% in Singapore	Sig.
Child Gender			
Male	169 (50.45)	49.30%	ns
Female	166 (49.55)	50.70%	
Monthly Household Income			
<SGD2000 (USD1450)	45 (14.38)	14.50%	$p < 0.001$
SGD2000-3999 (USD1450–2900)	92 (29.39)	21.20%	
\geq SGD4000 (USD2901)	176 (56.23)	64.30%	
Ethnicity			
Chinese	198 (60.00)	74.30%	$p < 0.001$
Malay	84 (25.45)	13.40%	
Indian	48 (14.55)	9.00%	
Maternal education			
\leq High School	84 (25.45)	57.60%	$p < 0.001$
College and above	246 (74.55)	42.40%	

Table 2 Descriptive data of the ASQ-3 and Bayley-III at 24 months.

	n	Mean (SD)	Minimum	Maximum
ASQ-3 24-Month				
Communication	332	47.56 (14.06)	0	60
Gross Motor	334	53.90 (8.93)	0	60
Fine Motor	333	45.71 (9.26)	15	60
Problem-Solving	330	43.94 (12.50)	5	60
Personal Social	335	43.16 (11.51)	10	60
Total ASQ-3 score	327	234.76 (40.37)	90	300
Bayley-III				
Cognitive	335	100.25 (14.59)	55	145
Language	332	100.22 (16.11)	47	141
Motor	331	100.35 (14.59)	46	139

cognitive, and motor domains, respectively, while 20 (6.08%) were identified as “delayed” in one or more Bayley-III domains.

3.3. Comparison of three sets of cut-off ASQ-3 scores

Table 3 describes three sets of ASQ-3 cut-off scores, along with sensitivity, specificity, NPV, PPV, and overall agreement across different domains, using Bayley-III scores <70 as the convergent measure. Using Criterion-I (Mean–1SD) cut-off scores on the ASQ-3, 11%–21% of children were identified as “at risk.” Sensitivity ranged from 14% to 33%, specificity from 82% to 91%, NPV from 69% to 74%, and PPV from 38% to 59%, with overall agreement from 64% to 71%.

Following Criterion-II (Mean–2SD) to generate ASQ-3 cut-off scores, 3–6% were identified as “at risk,” with sensitivity ranging from 20 to 47%, specificity from 94 to 98%, NPV from 95 to 97%, and PPV from 19 to 53%, with overall agreement from 90 to 94%.

Using Criterion III, to generate ROC for identification of cut-off scores with optimal agreement with the Bayley-III, 6–30% of children were identified as “at risk” on the ASQ-3, with a sensitivity of 20–74%, specificity from 72 to 94%, NPVs from 95 to 98%, and PPVs from 11 to 19%, with overall agreement from 72 to 90%.

3.4. Correlations between the ASQ-3 and Bayley-III

Significant correlations were seen between ASQ-3 domains and all five Bayley-III scales/sub-scales, as shown in Table 4.

Convergent validity. There were moderate to strong correlations between the ASQ-3 communication and Bayley-III language scales ($r = 0.44$ in RL, $r = 0.59$ in EL, $p < 0.01$). A moderate correlation was seen between ASQ-3 communication and Bayley-III cognitive domains ($r = 0.33$, $p < 0.01$) and between ASQ-3 personal social and Bayley-III EL domains ($r = 0.37$, $p < 0.01$). While the ASQ-3 GM and Bayley-III GM domains showed a moderate correlation ($r = 0.36$, $p < 0.01$), the ASQ-3 FM and Bayley-III FM domains showed a significant but much lower correlation ($r = 0.13$, $p < 0.05$). A low correlation coefficient was also seen for the ASQ-3 problem-solving and Bayley-III cognitive ($r = 0.22$, $p < 0.01$) domains.

Divergent validity. Correlation was much lower between conceptually divergent ASQ-3 and Bayley-III domains, such as ASQ-3 communication and Bayley-III GM ($r = 0.21$, $p < 0.01$) and FM ($r = 0.17$, $p < 0.01$) scales in

Table 3 ASQ-3 24-month cut-off score, number and percentage of children identified as “at-risk”, specificity, sensitivity PPV, NPV and overall agreement rate following three different cut-off criteria (n = 335).

ASQ-3 Cut-off Rule	ASQ-3 domain	ASQ-3 cut-off point	Risk of delay n (%)	Sensitivity (%)	Specificity (%)	NPV (%)	PPV (%)	Overall agreement (%)
Criterion-I: Mean – 1 standard deviation on ASQ-3	Communication	33.50	49 (14.76)	27.88	91.11	73.21	59.18	71.12
	Gross Motor	44.97	37 (11.08)	14.29	90.27	69.39	40.54	66.16
	Fine Motor	36.45	52 (15.62)	19.05	85.78	69.42	38.46	64.55
	Problem-Solving	31.44	69 (20.91)	26.67	81.53	70.16	40.58	63.91
	Personal-Social	31.64	64 (19.10)	33.02	87.17	73.51	54.69	69.88
	Total Score	194.39	56 (17.13)	29.81	88.64	72.76	55.36	69.75
Criterion-II: Mean – 2 standard deviation on ASQ-3	Communication	19.44	17 (5.12)	47.37	97.39	96.76	52.94	94.48
	Gross Motor	36.04	21 (6.29)	20.00	94.48	94.79	19.05	89.94
	Fine Motor	27.19	15 (4.50)	25.00	96.74	95.19	33.33	92.35
	Problem-Solving	18.94	11 (3.33)	20.00	97.70	94.89	36.36	92.90
	Personal-Social	20.13	18 (5.37)	30.00	96.12	95.50	33.33	92.10
	Total Score	154.02	12 (3.67)	21.05	97.35	95.15	33.33	92.83
Criterion-III: ROC optimal cut-offs of ASQ-3 using Bayley as the criterion	Communication	41.00	87 (26.20)	68.42	76.55	97.51	15.29	76.07
	Gross Motor	37.50	21 (6.29)	20.00	94.48	94.79	19.05	89.94
	Fine Motor	38.00	52 (15.62)	40.00	85.67	95.64	15.38	82.87
	Problem-Solving	35.50	92 (27.88)	50.00	73.36	95.71	10.99	71.91
	Personal-Social	38.75	100 (29.85)	70.00	72.49	97.39	14.14	72.34
	Total Score	211.50	85 (25.99)	73.68	76.49	97.88	16.47	76.32

Note: NPV = negative predictive values; PPV = positive predictive values.

Table 4 Correlations between the ASQ-3 at 24 months and Bayley-III (n = 335) (for consideration).

	Bayley Receptive Language	Bayley Expressive Language	Bayley Cognitive	Bayley Gross Motor	Bayley Fine Motor
ASQ-3 Communication	0.44**	0.59**	0.33**	0.21**	0.17**
ASQ-3 Gross Motor	0.11*	0.18**	0.18**	0.36**	0.18**
ASQ-3 Fine Motor	0.26**	0.22**	0.24**	0.17**	0.13*
ASQ-3 Problem-Solving	0.27**	0.23**	0.22**	0.18**	0.12*
ASQ-3 Personal Social	0.29**	0.37**	0.28**	0.16**	0.16**

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

contrast to the convergent domains. Similarly, ASQ-3 problem-solving and personal-social domains showed lower correlations with the GM and FM scales on Bayley-III. However, surprisingly, the ASQ-3 FM domain had a higher correlation with Bayley-III RL ($r = 0.26$, $p < 0.01$), EL ($r = 0.22$, $p < 0.01$), and cognitive ($r = 0.24$, $p < 0.01$) domains than with the convergent GM domain on Bayley-III ($r = 0.17$, $p < 0.01$).

4. Discussion

Our prospective study established local normative cut-off scores for ASQ-3 at 24 months of age and evaluated its validity as a screening tool using Bayley-III as the convergent measure. ASQ-3 cut-off scores based on Criterion-I seemed to be the most promising given the high specificity (82–91%) and NPV (69–74%) with moderate to strong correlations seen between ASQ-3 communication and Bayley-III receptive and expressive language scores.

4.1. Recommendations for ASQ-3 cut-off points

Our previous study¹⁴ reported significant differences between ASQ-3 scores in Singapore and the U.S. and suggested establishing local cut-offs. In the current study, the 24-month ASQ-3 local cut-off scores were derived from three different criteria (Table 3) along with identification of children “at-risk” for developmental delay. Based on Criterion-I, local ASQ-3 cut-off points were set at 1SD below the mean, resulting in 11–21% of the children at two years of age being identified as “at-risk” of delay across the five ASQ-3 domains. These rates are similar to the 18–20% figures reported by Sheldrick²¹ and Veldhuizen²² from low-risk primary care or general populations. The Criterion-II and Criterion-III ASQ-3 cut-off points had lower identification rates in certain domains, such as 3% in problem-solving based on Criterion-II and 5% in gross motor skills based on Criterion-III. As a screening measure, such low identification rates may reflect false-negative results and result in delayed referral/intervention.

ASQ-3 cut-off scores based on Criterion-I seemed to be the most promising given the high specificity (82–91%) and NPV (69–74%). The use of 1SD as a cut-off score at 24 months was also recommended by Lamsal et al.²³ as having better test characteristics as a screening tool compared to 2SD below the mean cut-off (Criterion II).

4.2. Sensitivity and specificity of ASQ-3 at 24 months

With all three criteria, we saw higher specificity (72–99%) and NPVs (69–98%) than sensitivity (19–74%) and PPV (11–59%). Specifically, all three ASQ-3 cut-off scores showed lower than expected (70%) sensitivity values, especially in the motor domains (GM and FM) (19–40%). This is similar to recently reported findings²¹ that screening tools, including the ASQ-3, offered adequate specificity but modest sensitivity for detecting developmental delays among young children. The authors also concluded that there was a trade-off between sensitivity and specificity for developmental screening tools. Specificity assesses how well a test would correctly identify children without developmental delay, and sensitivity estimates how well a test would correctly identify children with developmental delay. As a screening measure, sensitivity may be traded off for specificity as a test with low specificity would yield a large number of false-positive results with the ensuing unnecessary referrals and diagnostic assessments. This would limit the upscaling and generalizability of the screening test in low-risk populations.

Across all ASQ-3 domains as well as for the total ASQ scores, high NPV and specificity were seen in both our current study of low-risk children and our previous study of high-risk preterm VLBW survivors.¹³ This implies that a “typical development” screening result on the ASQ-3 could be used to provide a degree of reassurance to parents and avoid unnecessary caregiver anxiety, over-referrals, and expensive assessments. Anticipatory guidance should continue to be provided about appropriate preventive care and stimulation, and health care professionals should continue ongoing developmental surveillance and recommended age-appropriate developmental screening.²⁴ These measures would also provide a valuable opportunity build a positive relationship with families and enable family-centered care.

Our findings have also been corroborated by a national Canadian study,²³ which supports the use of ASQ-3 in identifying children not at risk of a neurodevelopmental disorder; and similarly other studies have shown usefulness of the ASQ in identifying children without a developmental delay.¹⁹

However, the ASQ-3’s sensitivity in the current study was lower (17%–74%) than the high sensitivity (84%) noted in our previous study¹³ in preterm survivors. One possible reason

for this was the low prevalence of developmental delay (2.5–6%) in our current low-risk cohort. The AAP reports a sensitivity range of 70–90% for the ASQ-3 based on a mixture of population and study types. However, a meta-analysis assessing the effectiveness of universal developmental screening in primary care settings included only studies conducted in low-risk populations and used standardized diagnostic evaluation.²⁵ That meta-analysis identified only four studies meeting inclusion criteria; and, for the ASQ-3, it reported a median sensitivity of 55.0% (47.1%–66.7%) and a median specificity of 86.0% (38.6%–94.3%). Similar to this meta-analysis, our study was also conducted in a low-risk population and used a standardized diagnostic evaluation and showed comparable sensitivity and specificity.

4.3. Correlations between the ASQ-3 and Bayley-III

The moderate to strong correlations seen between ASQ-3 communication and Bayley-III receptive and expressive language scores are clinically important, especially given Singapore's multilingual population. Language delay is among the most prevalent areas of developmental delay in young children and is reported in 5–12% of children between two and five years of age.²⁶ Communication delays have a negative effect on long-term academic, psychological, and social development. The significant correlations seen between the ASQ-3 communication with Bayley-III receptive/expressive language domains suggest that the ASQ-3 domain scores may help clinicians identify individual areas of delay and provide specifically targeted interventions.

The lower-than-expected correlation coefficients between ASQ-3 and Bayley-III raised concerns about the convergent validity of some of the ASQ-3 domain scores, especially in the ASQ-3 FM and problem-solving domains, where correlations with the corresponding Bayley-III scores were the lowest (0.12–0.21). Similar findings of significant but weak correlations between the two measures were reported in previous studies from China²⁷ and Chile.²⁸

4.4. Strengths of the current study

The study aimed at validating the ASQ-3 in a large low-risk cohort and it has the potential to answer questions about the ASQ-3's usefulness as a universal screening tool in the general population. Our study adds to the limited data available on using a developmental screening measure in low-risk populations, with corroboration provided by a standardized, diagnostic evaluation.

4.5. Limitations and directions for future research

The validity of the convergent Bayley-III measure requires further investigation as its use resulted in a relatively low identified prevalence of developmental delay (6%), and it showed weak correlations with the ASQ-3 in some domains. The Bayley-III has been normed on a non-Asian population, and thus there may be aspects of cultural sensitivity that were not taken into consideration in the current study. This

may be a possible reason for the weak correlations in the non-language ASQ domains with the Bayley-III.

While our study has examined validity of the ASQ-3 in Singapore's multi-ethnic context, the validation of the Bayley-III as a diagnostic measure requires further research. Another direction for future research is to evaluate the relation of the ASQ-3 to longer-term outcomes (e.g., in school age).

5. Conclusions

The current study reported and compared three sets of ASQ-3 cut-off points based on different methods and criteria. Recommendations for selecting and using appropriate ASQ-3 cut-off points in clinical practice were made accordingly. The high specificity of the ASQ-3 reflects its value in identifying typically developing children not at risk of developmental delay. It also helps to provide reassurance to parents about their child's typical development, avoiding unnecessary referrals and expensive assessments, and minimizing the burden on family community pediatricians and developmental specialists.

Conflict of interest statement

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

Acknowledgements

KMG is supported by the UK Medical Research Council (MC_UU_12011/4), the National Institute for Health Research (NIHR Senior Investigator (NF-SI-0515-10042) and NIHR Southampton Biomedical Research Centre (IS-BRC-1215-20004)), the European Union (Erasmus + Programme ImpENSA 598488-EPP-1-2018-1-DE-EPPKA2-CBHE-JP) and the British Heart Foundation (RG/15/17/3174, SP/F/21/150013).

References

1. Black MM, Walker SP, Fernald LCH, Andersen CT, DiGirolamo AM, Lu C, et al. Early childhood development coming of age: science through the life course. *Lancet* 2017;389:77–90.
2. McLean K, Goldfeld S, Molloy C, Wake M, Oberklaid F. *Screening and surveillance in early childhood health: rapid review of evidence for effectiveness and efficiency of models*. Ultimo, NSW, Australia: The Sax Institute; 2014.
3. Moore TG, McDonald M. *Acting early, changing lives: how prevention and early action saves money and improves well-being*. Victoria: Prepared for The Benevolent Society, Centre for Community Child Health at The Murdoch Childrens Research Institute and The Royal Children's Hospital. Parkville; 2013.
4. King TM, Glascoe FP. Developmental surveillance of infants and young children in pediatric primary care. *Curr Opin Pediatr* 2003;15:624–9.
5. Bellman M, Byrne O, Sege R. Developmental assessment of children. *BMJ* 2013;346:e8687.
6. Boyle CA, Boulet S, Schieve La, Cohen RA, Blumberg SJ, Yearigi-Allsopp M, et al. Trends in the prevalence of developmental

- disabilities in US Children; 1997–2008. *Pediatrics* 2011;127:1034–42.
7. Zablotsky B, Black LI, Maenner MJ, Schieve LA, Blumberg SJ. Estimated prevalence of autism and other developmental disabilities following questionnaire changes in the 2014 National Health Interview Survey. *Natl Health Stat Report* 2015;87:1–20.
 8. 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* 2006;118:405–20.
 9. Thomas RE, Spragins W, Mazloum G, Cronkhite M, Maru G. Rates of detection of developmental problems at the 18-month well-baby visit by family physicians' using four evidence-based screening tools compared to usual care: a randomized controlled trial. *Child Care Health Dev* 2016;42:382–93.
 10. Glascoe FP, Foster EM, Wolraich ML. An economic analysis of developmental detection methods. *Pediatrics* 1997;99:830–7.
 11. Squires J, Bricker DD. *Ages & stages questionnaires*. 3rd ed. Baltimore, MD: Paul H. Brookes; 2009.
 12. Shonkoff JP. Building a new biodevelopmental framework to guide the future of early childhood policy. *Child Dev* 2010;81:357–67.
 13. Agarwal PK, Shi L, Daniel LM, Yang PH, Khoo PC, Quek BH, et al. Prospective evaluation of the Ages and Stages Questionnaire 3rd Edition in very-low-birthweight infants. *Dev Med Child Neurol* 2017;59:484–9.
 14. Agarwal PK, Xie H, Sathyapalan Rema AS, Rajadurai VS, Lim SB, Meaney M, et al. Evaluation of the ages and stages questionnaire (ASQ 3) as a developmental screener at 9, 18, and 24 months. *Early Hum Dev* 2020;147:105081.
 15. Soh SE, Tint MT, Gluckman PD, Godfrey KM, Rifkin-Graboi A, Chan YH, et al. Cohort profile: Growing up in Singapore towards Healthy Outcomes (GUSTO) birth cohort study. *Int J Epidemiol* 2014;43:1401–9.
 16. Bayley N. *Bayley scales of infant and toddler development*. San Antonio, TX: Psychological Corporation; 2006.
 17. Steenis LJ, Verhoeven M, Hessen DJ, van Baar AL. Performance of Dutch children on the Bayley III: a comparison study of US and Dutch norms. *PLoS One* 2015;10:e0132871.
 18. Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285–93.
 19. Steenis LJP, 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;91:217–25.
 20. (GHS) 2015 *Online archive of general household survey*. Singapore: Department of Statistics of Ministry of Trade and Industry; 2015. Available at, https://www.singstat.gov.sg/publications/cop2010/census10_stat_release1.
 21. Sheldrick RC, Marakovitz S, Garfinkel D, Carter AS, Perrin EC. Comparative accuracy of developmental screening questionnaires. *JAMA Pediatr* 2020;174:366–74.
 22. 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;15:231–7.
 23. Lamsal R, Dutton DJ, Zwicker JD. Using the ages and stages questionnaire in the general population as a measure for identifying children not at risk of a neurodevelopmental disorder. *BMC Pediatr* 2018;18:122.
 24. Lipkin PH, Macias MM. Council on Children with Disabilities, Section on Developmental and Behavioral Pediatrics. Promoting optimal development: identifying infants and young children with developmental disorders through developmental surveillance and screening. *Pediatrics* 2020;145:e20193449.
 25. Warren R, Kenny M, Fitzpatrick-Lewis D. *Screening and treatment for developmental delay in early childhood (ages 1–4): systematic review*. Hamilton, Ontario: McMaster University; 2014.
 26. Berkman ND, Wallace I, Watson L. *Screening for speech and language delays and disorders in children age 5 years or younger: a systematic review for the U.S. Preventive Services Task Force*. Rockville (MD): Agency for Healthcare Research and Quality (US); 2015. Available at, <https://www.ncbi.nlm.nih.gov/kkh.remotexts.co/books/NBK305676/>.
 27. Yue A, Jiang Q, Wang B, Abbey C, Medina A, Shi Y, et al. Concurrent validity of the ages and stages questionnaire and the Bayley scales of infant development III in China. *PLoS One* 2019;14:e0221675.
 28. Schonhaut L, Armijo I, Schönstedt M, Alvarez J, Cordero M. Validity of the ages and stages questionnaires in term and preterm infants. *Pediatrics* 2013;131:e1468–74.