



NIH PUBLIC ACCESS

Author Manuscript

Eur J Paediatr Neurol. Author manuscript; available in PMC 2015 March 01.

Published in final edited form as:

Eur J Paediatr Neurol. 2014 March ; 18(2): 223–230. doi:10.1016/j.ejpn.2013.11.011.

Validity of US norms for the Bayley Scales of Infant Development-III in Malawian children

Elizabeth A Cromwell, MPH^a, Queen Dube, MD^{b,c}, Stephen R Cole, PhD^a, Chawanangwa Chirambo, BS^b, Anna E Dow, MSPH, PhD^a, Robert S Heyderman, MD^b, and Annelies Van Rie, MD, PhD^{a,*}

^aDepartment of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill

^bMalawi-Liverpool-Wellcome Trust Clinical Research Programme, University of Malawi College of Medicine, Blantyre, Malawi

^cDepartment of Pediatrics & Child Health, Queen Elizabeth Central Hospital, Blantyre, Malawi

Abstract

Objective—Most psychometric tests originate from Europe and North America and have not been validated in other populations. We assessed the validity of United States (US)-based norms for the Bayley Scales of Infant and Toddler Development-III (BSID-III), a neurodevelopmental tool developed for and commonly used in the US, in Malawian children.

Methods—We constructed BSID-III norms for cognitive, fine motor (FM), gross motor (GM), expressive communication (EC) and receptive communication (RC) subtests using 5 173 tests scores in 167 healthy Malawian children. Norms were generated using Generalized Additive Models for location, scale and shape, with age modeled continuously. Standard z-scores were used to classify neurodevelopmental delay. Weighted kappa statistics were used to compare the classification of neurological development using US-based and Malawian norms.

Results—For all subtests, the mean raw scores in Malawian children were higher than the US normative scores at younger ages (approximately <6 months) after which the mean curves crossed and the US normative mean exceeded that of the Malawian sample and the age at which the curves crossed differed by subtest. Weighted kappa statistics for agreement between US and Malawian norms were 0.45 for cognitive, 0.48 for FM, 0.57 for GM, 0.50 for EC, and 0.44 for RC.

Conclusion—We demonstrate that population reference curves for the BSID-III differ depending on the origin of the population. Reliance on US norm-based standardized scores

© 2013 European Paediatric Neurology Society. Published by Elsevier Ltd. All rights reserved.

***Corresponding Author:** Dr. Annelies Van Rie, University of North Carolina, 2107 McGavran-Greenberg, Chapel Hill, NC 27599, 919-966-1420, vanrie@email.unc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Financial Disclosure: All authors have no financial relationships relevant to this article.

Conflict of Interest: All authors have no conflict of interest to disclose.

resulted in misclassification of the neurological development of Malawian children, with the greatest potential for bias in the measurement of cognitive and language skills.

Keywords

Child development; testing norms; cross-cultural testing bias

Introduction

It has been estimated that more than 200 million children under 5 years of age are not reaching their full potential for growth, cognition, or socio-emotional development due to risk factors for neurological delay.¹ In the African setting, poverty, human immunodeficiency virus (HIV) infection, malaria and malnutrition have been shown to adversely affect neurological development.¹⁻⁵ While it is critical to assess whether children develop appropriately, few national statistics from developing countries exist and neurodevelopmental delay in these settings remains understudied. In addition to the paucity of data on pediatric neurological development in Africa, most extant data were collected using assessments developed in Europe or North America. This is, at least in part, due to the scarcity of psychometric tools to measure neurological development specifically designed for settings outside of Europe and North America.

Diagnosis of a child's neurological development requires comparison to a normative reference population by transforming raw scores to percentile ranks or standardized ("scaled") scores.⁶ Normative samples are usually cross-sectional, drawn from healthy children who represent the target population. For example, the *Bayley Scales of Infant and Toddler Development, Third Edition* (BSID-III), is widely used in international child development research. The normative sample for the BSID-III included 1 700 children stratified by age, sex, parental education, race and geographic region in the United States (US).⁷ Data from the normative sample were then used to construct norms that represent the distribution of test performance in the US population. Test norms are necessary to identify delay; without the ability to compare a child's performance to what is considered "normal", raw scores have limited value for research or clinical practice.

Several authors have raised concerns about the use of a psychometric tool in a population other than the one in which the tool was developed.⁸⁻¹⁰ To address possible cross-cultural bias when using developmental assessments in populations other than those for whom norms were developed, three approaches have been commonly used. First, an entirely new test can be developed and normed for a specific population.¹¹ Successful generation of a new test involves an inter-disciplinary research team, an adequate representative sample for testing items and test cohesion, and the concurrent development of norms or standards that represent typical development.¹² For example, the Kilifi Developmental Inventory is a new tool designed to monitor and describe the development of at-risk children in resource-limited settings in Kenya.¹¹ A new test ensures a culturally appropriate psychometric tool but its development is resource-intensive and prohibits use beyond its target population, thus limiting comparability of findings with other tests and across populations.

An alternative to the development of a new test is the adaptation of existing tools for use in new populations. Published guidelines have been developed for this process, which typically involves translation of test materials, modification of test items inappropriate for the local context to preserve consistency with the constructs measured, followed by a process of iterative adaptation and testing of the assessment tool.^{12–14} For example, Nampijja *et al* successfully adapted several Western measures to assess cognition in 5-year-old semi-urban Ugandan children.¹⁵ Gladstone *et al* employed a hybrid approach in which existing test items were modified for local context and combined with new test items designed specifically for the population under study.¹⁶ While an adapted test improves cultural appropriateness, it is also resource-intensive and does not permit comparability between populations. Furthermore, although adaptation of existing assessment tools may reduce bias in test items measuring specific constructs, there remains a risk of bias unless these adaptations are accompanied by creation of local norms as children in one setting may perform better, on average, than children in another due to cultural differences in child rearing and access to early education.⁸

Finally, tests developed in Europe and North America have been used in other populations in epidemiologic studies of specific exposures without adaptation by employing a healthy control group for comparison. For example, controls were enrolled in a study comparing the neurodevelopment of HIV-infected versus HIV-exposed uninfected children in the Democratic Republic of Congo.¹⁷ Healthy children who serve as a control group can be used to reduce norm-related bias when assessing group-level differences in developmental delay. Existing tests without adaptation can also be used to identify factors associated with neurodevelopmental scores within one population, as was done in a study of the effect of stunting and wasting on the neurodevelopment of Tanzanian children.¹⁸ However, while use of controls allows comparisons of scores between groups, this approach does not allow the unbiased diagnosis of delay for individual children.

We present an alternative approach to the challenge of neurodevelopmental assessment of children in resource-poor countries by developing new population norms. We applied this method to the BSID-III using data collected in healthy Malawian children age 10 weeks to 30 months and compared the classification of neurological delay using the Malawian and US norms.

Materials and Methods

Study site and selection of participants

From March 2008 to December 2009, 167 healthy children were enrolled as controls for a cohort study of HIV infection and neurological development. HIV-negative mothers age 15 years without a history of alcohol or substance abuse were randomly selected among women attending prenatal care visits at two primary care facilities in Blantyre, Malawi. Their infants were enrolled at age 10 and 14 weeks if born without congenital abnormalities and free of severe disease at enrollment. Children were confirmed HIV-negative via polymerase chain reaction (Roche Amplicor) at enrollment and screened for HIV infection by rapid test (Determine and Unigold Rapid Tests) at age 30 months, or when follow-up ended. A questionnaire was administered to the mother or primary caregiver to measure

proxies for wealth (e.g. electricity, telephone, sanitation in the home) to assess whether the distribution of socio-economic status among the study participants was comparable to that reported in the Malawian 2010 Demographic and Health Survey.¹⁹

Administration of BSID-III

The BSID-III⁷ is a psychometric tool designed to measure neurological development in children one month to 42 months of age. Five subtests of the BSID-III were administered to each child at 10 weeks, 14 weeks, and then 6, 9, 12, 15, 18, 24, and 30 months of age: receptive communication, expressive communication, fine motor, gross motor and cognitive. Tests were implemented by trained study staff according to the BSID-III guidelines, using the child's age (in months) to determine the starting test item for each subtest and applying the reversal and discontinue rules.²⁰ Where necessary, instructions and test materials were translated into Chichewa, with approval from Pearson Education, Inc. In addition, minor modifications to the test scoring were made to ensure cultural appropriateness. For example, for the BSID-III item depicting an apple a correct score was assigned when the child identified the "apple" as a "tomato" as apples are not familiar to most Malawian children.

Calculation of BSID-III raw and scaled scores

The raw score was calculated as the number of test items that preceded the starting item plus the number of items completed by the child. Test scores were recorded on standard BSID-III record forms. The study manager verified the calculation of the raw scores before data entry. Raw scores were double-entered using the BSID-III Scoring Assistant; discrepant entries were resolved by consulting the paper record. Four subtests were not included due to discrepant score entries that could not be resolved. The Scoring Assistant automatically converted raw scores into scaled scores based on the US norms.

Development of Malawian *BSID-III* norms

Standardized scores from normative test data can be calculated with a variety of statistical approaches. The simplest is to estimate percentiles; more complex approaches include modeling the distribution of raw scores parametrically as a function of a variable used for comparison, such as age, and then deriving z-scores.^{21,22} We used Generalized Additive Models for Location, Scale and Shape (GAMLSS) as described by Rigby and Stasinopolos to generate normative reference curves and standardized scores from the data collected longitudinally in 167 Malawian children.^{23–25} For each subtest, we first identified the optimal transformation of age to model the raw BSID-III score and identified the degrees of freedom necessary to model the median, variance, scale and shape parameters. Normative curves were developed in R using the model fitting functions provided by the GAMLSS package and the `centiles.pred` function to compare fitted percentiles to observed percentiles and generate standardized z-scores.²⁶

To determine if there was substantial correlation due to repeated measures (a consequence of our longitudinal approach to data collection), we conducted a sensitivity analysis using a non-parametric bootstrap to compare standard errors of a regression model with age and a generalized estimating equation with age controlling for cluster effects by child. Both models regressed age with a quadratic term and used the raw score as the outcome variable.

A priori, we decided that a 10% change in the standard error would suggest correlation among repeated measurements. We only present data from GAMLSS models without controlling for repeat measurements as none of the comparisons among the five subtests exceeded this threshold.

Comparison of US and Malawian norms

We visually compared the shape of the mean raw scores from the US and Malawi norms to determine if the trajectory of scores by age differed. We used the normative curves produced by the GAMLSS models to plot the shape of the Malawian mean raw score for each subtest. We used the raw score conversion charts included in the BSID-III administration materials to estimate the mean for the 48 BSID-III age categories (as the US normative data are not publicly available) to plot the mean raw score by age for the US norms.²⁷

We also classified neurological development for each individual child according to both the US and Malawian norms as described in Table 1. For the Malawian norms, raw scores from study participants were directly converted by the GAMLSS program into z-scores. For the US norms, US scaled scores were automatically calculated by the BSID-III Scoring Assistant and scaled scores are normally distributed with a mean of 10 (standard deviation (SD) of 3) and range 1–19. Given the ordinal nature of neurological classification based on the overall BSID-III score for each subtest, we estimated the agreement between the US and Malawian norms with weighted kappa statistics using linear weights in SAS version 9.2.

Ethical Considerations

Parents or guardians of all participants gave written informed consent before study enrollment. Approval to conduct the research was granted by the University of Malawi College of Medicine Research and Ethics Committee and the University of North Carolina Institutional Review Board. Test results were not communicated as the scores could not be interpreted until local norms were developed. Instead, all parents were given age-specific advice on how to stimulate their child's development independent of the child's performance.

Results

Study population characteristics

A total of 167 children contributed 5 173 unique BSID-III sub-tests. The majority (70%) completed 12 months follow-up, most (59%) completed 24 months follow-up, and 50% completed assessments at 30 months of age. Children who were lost to follow-up contributed to the analysis through their last study visit. Reasons for incomplete follow-up were relocation or withdrawal (n=73), death (n=6), and administrative censoring at end of study (n=7). Except for a higher proportion of mothers being < 30 years of age among children not completing the study (p=0.01), children who did and did not complete 30 months of follow up were similar with respect to gender (p=0.7642), household ownership of animals (p=0.3545), mother's marital status (p=0.6895) and level of mother's education (p=0.9682). Some children did not complete all five tests during a study visit resulting in a variable number of completed subtests: 1 027 for cognitive; 1 035 for expressive

communication; 1 037 for receptive communication; 1 036 fine motor and 1 038 gross motor. A total of four subtests were not included due to unresolvable data entry errors. Approximately 13% of study visits were missed (not including visits after a child was censored). Data from an individual child was included through 30 months or loss to follow-up.

Of the children enrolled, 81 were girls (48.5%) and 113 (67.7%) had at least one sibling in the household. Most primary caregivers were between 20–29 years of age (68.2%) and 79.6% were married. The prevalence of wealth indicators among households of study participants was similar to those reported by the 2010 Malawi Demographic and Health survey (Table 2).

Malawian *BSID-III* norms

Figure 1 presents the Malawian normative curves, with the mean, plus and minus one and two standard deviations as a function of age for receptive communication, expressive communication, fine motor, gross motor and cognitive development. Similar to widely used charts for weight and height, these charts allow the plotting of individual *BSID-III* raw scores to classify children's development as normal (between mean and plus or minus 1 SD), mild delay (between minus 1 and minus 2 SD), severely delayed (> minus 2 SD), above normal (between mean and plus 1 SD) or advanced (> plus 2 SD above mean) for each of the five subtests.

Comparison of US and Malawian norms for five *BSID-III* subtests

In Figure 2, we compare the mean of the US norms with the mean of the Malawian norms and show that the distribution of neurological performance in our normative Malawian population is not merely shifted above or below the mean performance measured in the US. Rather, the mean raw score among the normative Malawian population was higher before age 6 months for the language subtests and before 12 months of age for the cognitive subset, after which the mean curves crossed and the US normative mean exceeded that of our Malawian sample.

Classification of 167 control children by US and Malawian norms

Approximately 70% of all tests administered to the participating children were classified as normal by both norms (Table 3). Weighted kappa coefficients for the five subtests ranged from 0.44 to 0.57 suggesting moderate agreement between the classification using the US and Malawi norms. Cognitive and receptive communication subtests had the lowest kappa scores (0.45 and 0.44, respectively) classification of gross motor demonstrated the highest level of agreement (0.57). Using the US-based norms, the following number of tests would have been misclassified: fine motor, 301 (29.1%); gross motor, 265 (25.5%); receptive communication, 319 (30.8%); expressive communication, 261 (25.2%); and cognitive, 372 (36.2%).

Discussion

Despite the interest in examining effects of biological and environmental factors on cognitive functioning of children in low and middle income countries, many psychometric tests originate from Europe and North America and have not been validated in other populations. We present the creation of locally-normed population reference curves for a neurodevelopmental tool (BSID-III) developed for and commonly used in the US. Our data show that the BSID-III norms developed from test scores of Malawian children are not equivalent to the US-based norms and that reliance on US norm-based standardized scores may result in misclassification of the neurological development of Malawian children. The relationship between the two norms was not uniform but differed across age groups and differed for each subset.

These results are unsurprising. Scores on cognitive development tests have been shown to be higher among populations with greater access to education, urbanization and economic security.²⁸⁻³² Compared to the United States, early childhood in the African setting is characterized by vastly different childcare practices and different exposure to educational games, books, toys, and multimedia on television and personal computers. These practices likely result in cultural differences in neurological development, especially in the domains of language and cognition.

If the normative curves from our sample were the same shape as the US curves, one could simply rescale (either upwards or downwards) US norms for different populations. We demonstrate that the shape of test performance differs between Malawian and US normative populations, with Malawian children having higher raw scores in the first months of life and US children having higher scores after the first year of life, with increasingly diverging mean curves for all subtests except fine motor skills. The results suggest that the US norms may be useful in the first year of life. However, even when the normative means are similar, as was the case for the fine motor development, the variability in scores resulted in differences in neurodevelopmental classification (weighted kappa= 0.48), suggesting that different norms may be needed for all age groups.

Our study has several limitations. First, we would ideally have employed the same statistical method used to create the US norms for the BSID-III. Unfortunately, this information is not publicly accessible. Instead, we used GAMLSS models, the methodology used by the World Health Organization to develop norms for child growth curves.³³ GAMLSS is a more flexible technique compared to traditional methods which assume that raw scores are normally distributed. Second, we used longitudinal assessments in a cohort of children. Most normative samples are cross-sectional to avoid correlation due to repeated measures. In a sensitivity analysis, we did not find evidence of correlation (data not shown). Third, our normative sample size of 167 children may seem small compared the cross-sectional sample of 1 700 US children sampled from 17 age groups. Our longitudinal approach did however result in 1 030 assessments in 167 children prospectively followed from age 10 weeks to 30 months. Furthermore, we assessed more children per age group (n=167) than the US normative sample (n=100). Finally, we did not enroll a random sample of healthy children. Instead, we randomly selected from HIV negative mothers who had participated in an HIV

testing program for prevention of mother-to-child transmission. As such, the children enrolled may not have been fully representative of Malawian children, but are likely representative of the population for which these norms were initially developed. These new norms serve to facilitate the comparison of neurological development in a cohort of HIV-infected and HIV-exposed children with the neurological development of healthy, HIV-uninfected and unexposed children from the same source population. A comparison of household characteristics among study subjects with demographic data from the Malawian DHS suggests our sample was representative of urban and semi-urban Malawian households. Finally, we did not administer a second neurological assessment tool to validate diagnosis of delay in our study population.

In conclusion, our findings demonstrate that reliance on US-based norms for the BSID-III in Malawian children results in misclassification of neurological developmental delay and that the cross-cultural bias introduced when using developmental assessments in populations with substantially different child rearing practice can be addressed by development of local norms. Future research should assess if norms developed for populations other than the US and Europe are valid within larger geographic regions, similar to the age-specific norms developed by the World Health Organization for weight and height.

Acknowledgments

Funding was provided by the United States National Institutes of Health/National Institute of Child Health and Human Development, via grant R01 HD053216. MLW is funded through a core program grant from the Wellcome Trust, UK. We acknowledge Jill Lebov, Caroline Hexdall and Daniel Lowe for their contributions to the study. We also thank the participants and their families for their participation.

Abbreviations

BSID-III	Bayley Scales of Infant and Toddler Development
EC	expressive communication
FM	fine motor
GAMLSS	Generalized Additive Model for Location Scale and Shape
GM	gross motor
HIV	Human Immunodeficiency Virus
RC	receptive communication
SD	standard deviation
US	United States

References

1. Grantham-McGregor S, Cheung YB, Cueto S, et al. Developmental potential in the first 5 years for children in developing countries. *Lancet*. 2007; 369(9555):60–70. [PubMed: 17208643]
2. Grantham-McGregor S, Baker-Henningham H. Review of the evidence linking protein and energy to mental development. *Public Health Nutrition*. 2005; 8(7A):1191–1201. [PubMed: 16277829]

3. Le Doare K, Bland R, Newell ML. Neurodevelopment in children born to HIV-infected mothers by infection and treatment status. *Pediatrics*. 2012; 130(5):e1326–e1344. [PubMed: 23118140]
4. Idro R, Marsh K, Chandy CJ. Cerebral malaria: Mechanisms of brain injury and strategies for improved neuro-cognitive outcome. *Pediatric Research*. 2010; 86(4):267–274. [PubMed: 20606600]
5. Grantham-McGregor SM, Ani CC. Undernutrition and mental development. Nestle Nutrition workshop series. *Clinical & Performance Programme*. 2001; 5:1–14. discussion 14–18. [PubMed: 11510433]
6. Cicchetti D. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*. 1994; 6(4):284–290.
7. Bayley, N. Bayley Scales of Infant and Toddler Development-Third Edition: Technical Manual. San Antonio, TX: Harcourt Assessment; 2006.
8. Geisinger KF. Cross-cultural normative assessment: translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*. 1994; 6(4):304–312.
9. Van de Vijver FJR, Hambleton RK. Translating tests: some practical guidelines. *European Psychologist*. 1996; 1(2)
10. Poortinga YH. Cultural bias in assessment: Historical and thematic issues. *European Journal of Psychological Assessment*. 1995; 11(3):140–146.
11. Abubakar A, Van De Vijver FJ, Mithwani S, et al. Assessing developmental outcomes in children from Kilifi, Kenya, following prophylaxis for seizures in cerebral malaria. *J Health Psychol*. 2007; 12(3):417–430. [PubMed: 17439993]
12. Fernald, LCH.; Kariger, P.; Engle, P.; Raikes, A. Examining early child development in low-income countries: A toolkit for the assessment of children in the first five years of life. Washington, DC: The World Bank; 2009.
13. Hambleton RK. Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment*. 1994; 10(3):229–244.
14. Hambleton, RK. Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In: Hambleton, RK.; Merenda, PF.; Spielberger, CD., editors. *Adapting educational and psychological tests for cross-cultural assessment*. New Jersey: Lawrence Erlbaum Associates Inc; 2005. p. 3-37.
15. Nampijja M, Apule B, Lule S, et al. Adaptation of Western measures of cognition for assessing 5-year-old semi-urban Ugandan children. *The British journal of educational psychology*. 2010; 80(Pt 1):15–30. [PubMed: 19594989]
16. Gladstone M, Lancaster GA, Umar E, et al. The Malawi Developmental Assessment Tool (MDAT): The creation, validation and reliability of a tool to assess child development in rural African settings. *PLoS Med*. 2010; 7(5)
17. Van Rie A, Mupuala A, Dow A. Impact of the HIV/AIDS epidemic on the neurodevelopment of preschool-aged children in Kinshasa, Democratic Republic of the Congo. *Pediatrics*. 2008; 122(1):e123–e128. [PubMed: 18595957]
18. McDonald CM, Manji KP, Kupka R, et al. Stunting and wasting are associated with poorer psychomotor and mental development in HIV-exposed Tanzanian infants. *The Journal of Nutrition*. 2013; 143(2):204–214. [PubMed: 23256148]
19. Malawi Demographic and Health Survey 2010. Zomba, Malawi and Claverton, Maryland: NSO and ICF Macro; 2011.
20. Bayley, N. Bayley Scales of Infant and Toddler Development Screening Test: Third Edition. San Antonio, TX: Harcourt Assessment Inc; 2006. p. 138
21. Zachary RA, Gorsuch RL. Continuous norming: implications for the WAIS-R. *Journal of clinical psychology*. 1985; 41(1):86–94. [PubMed: 3973045]
22. Zhu J, Chen H. Utility of inferential norming with smaller sample sizes. *Journal of Psychoeducational Assessment*. 2011; 29(6):570–580.
23. Stasinopoulos DM, Rigby RA. The GAMLSS. www.gamlss.org.
24. Rigby RA, Stasinopoulos DM. Automatic smoothing parameter selection in GAMLSS with an application to centile estimation. *Statistical Methods in Medical Research*. 2013 Feb.

25. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape (with discussion). *Appl Statist.* 2005; 54(3):507–554.
26. Stasinopoulos DM, Rigby RA. Generalized additive models for location scale and shape (GAMLSS) in R. *J Stat Software.* 2007; 23:1–46.
27. Bayley, N. *The Bayley Scales of Infant and Toddler Development: Administration manual.* San Antonio, TX: PsychCorp; 2006.
28. Ceci SJ. How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental psychology.* 1991; 27(5):703–722.
29. Ceci SJ, Williams WM. Schooling, Intelligence, and Income. *American Psychologist.* 1997; 52(10):1051–1058.
30. Hansen KT, Heckman JJ, Mullen KJ. The effect of schooling and ability on achievement test scores. *Journal of Econometrics.* 2004; 121(1–2):3.
31. Alcock KJ, Holding PA, Mung'ala-Odera V, Newton CRJC. Constructing tests of cognitive abilities for schooled and unschooled children. *Cross-Cultural Psychology.* 2008; 39(5):529–551.
32. Kaplan-Sanoff M, Parker S, Zuckerman B. Poverty and early childhood development: What do we know, and what should we do? *Infants & Young Children.* 1991; 4(1):68–76.
33. Borghi E, de Onis M, Garza C, et al. Construction of the World Health Organization child growth standards: selection of methods for attained growth curves. *Statistics in Medicine.* 2006; 25(2): 247–265. [PubMed: 16143968]

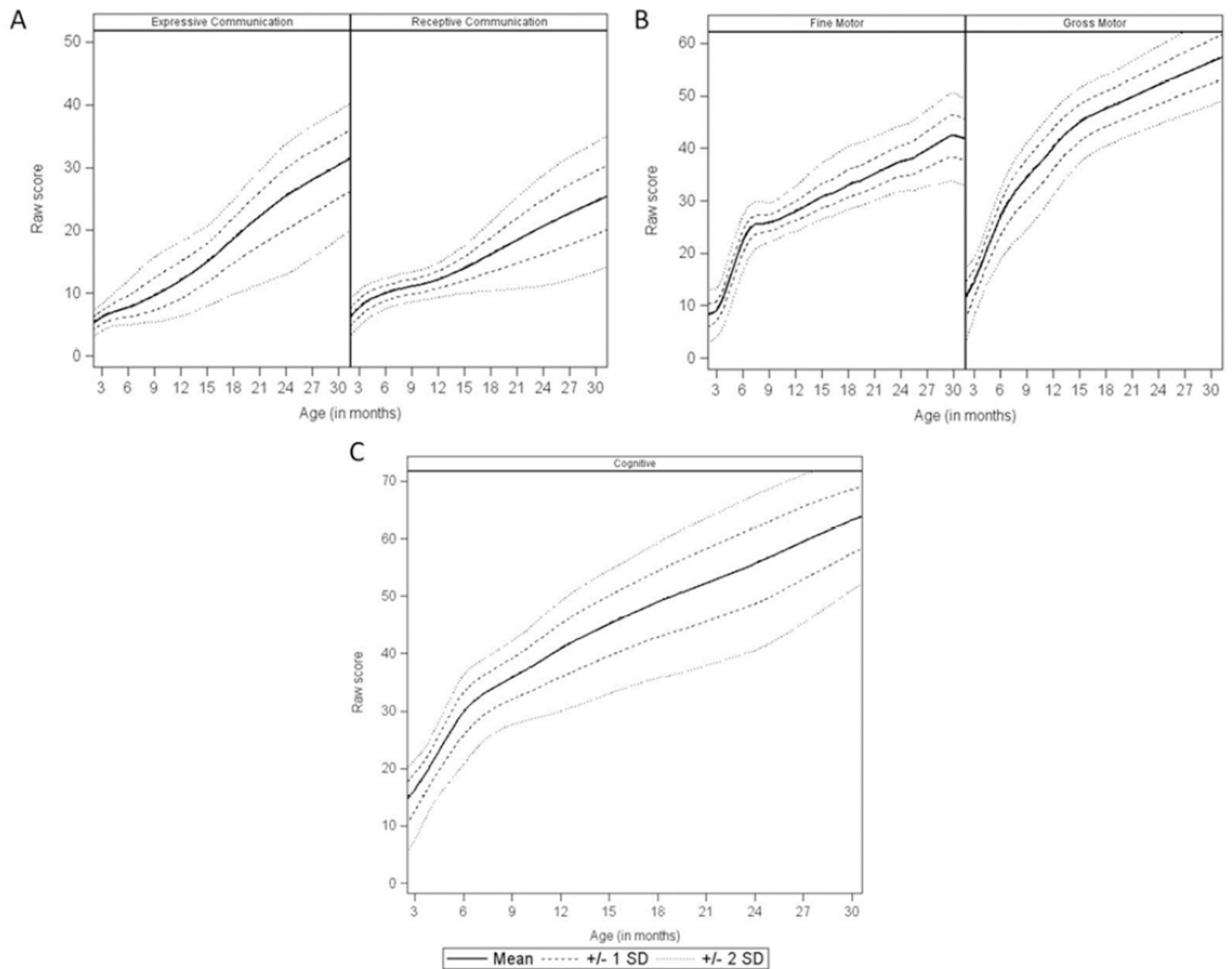


Figure 1. Charts for standard reference curves of raw scores for five BSID-III subtests from Malawian sample

Standard reference curves for (A) expressive and receptive communication, (B) fine and gross motor and (C) cognitive are presented with the mean (solid line), first and second standard deviations (dotted lines).

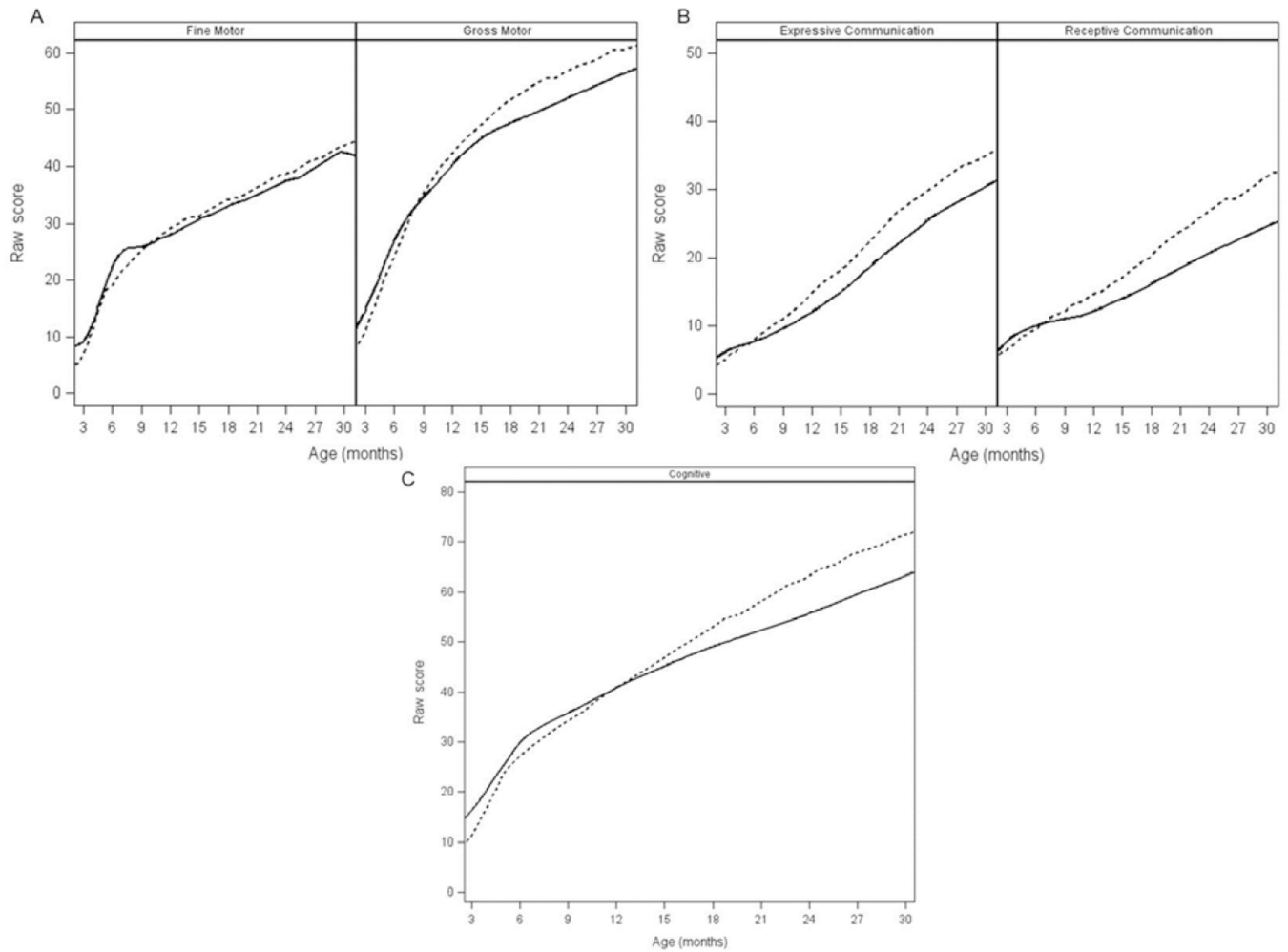


Figure 2. Comparison of mean raw score using the US and Malawian norms for five BSID-III subtests

The mean for the Malawian normative curve (solid line) is compared to an estimate of the mean US normative curve (dotted line) for (A) expressive and receptive communication, (B) fine and gross motor and (C) cognitive subtests.

Table 1

Classification of neurological development according to scores from US and Malawian norms

Level of delay	Range of US scaled scores	Range of Malawian z-scores
Severe delay	1–3	< -2
Mild delay	4–6	-2 raw score < -1
Normal	7–13	-1 raw score 1
Above normal	14–16	1 < raw score 2
Advanced	17–19	> 2

Table 2

Household characteristics of 167 Malawian children included in the normative sample compared to urban Malawi

Household Possessions	Number of households (%)	Malawi DHS19 % [‡]
Car	5 (3)	6.5
Motorbike	1 (1)	1.2
Bicycle	36 (24)	29.9
Television	39 (26)	34.2
Radio	118 (78)	70.3
Refrigerator	16 (11)	15.7
Flush toilet	11 (7)	9.4
Pit/bucket latrine	140 (93)	86.4
Electricity	56 (37)	34.7
Mobile phone	111 (75)	73.0
Farm animals	55 (68)	26.5

[‡]Prevalence estimates for urban Malawi as the two study sites were located in urban or peri-urban areas of Blantyre.

Table 3

Comparison of classification of neurodevelopment of 167 children based on Malawi and US norms

BSID-III subtest	Severely Delayed		Delayed		Normal		Above Normal		Advanced		Weighted Kappa κ (95% CI)
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	
<i>Cognitive</i> (n = 1027)											
Malawi Norms	22 (2.1)	146 (14.2)	706 (68.7)	130 (12.7)	23 (2.2)	0.45 (0.40, 0.50)					
US Norms	23 (2.2)	121 (11.8)	658 (64.1)	187 (18.2)	38 (3.7)						
<i>Expressive Communication</i> (n = 1035)											
Malawi Norms	28 (2.7)	137 (13.2)	695 (67.2)	155 (15.0)	20 (1.9)	0.50 (0.45, 0.55)					
US Norms	14 (1.4)	160 (15.5)	816 (78.8)	44 (4.3)	1 (0.1)						
<i>Receptive Communication</i> (n = 1037)											
Malawi Norms	31 (3.0)	112 (10.8)	752 (72.5)	115 (11.1)	27 (2.6)	0.44 (0.39, 0.49)					
US Norms	27 (2.6)	205 (19.8)	732 (70.6)	64 (6.2)	9 (0.9)						
<i>Fine Motor</i> (n = 1036)											
Malawi Norms	36 (3.5)	131 (12.6)	716 (69.1)	128 (12.4)	25 (2.4)	0.48 (0.43, 0.53)					
US Norms	9 (0.9)	41 (4.0)	780 (75.3)	176 (17.0)	30 (2.9)						
<i>Gross Motor</i> (n = 1038)											
Malawi Norms	30 (2.9)	121 (11.7)	715 (68.9)	149 (14.4)	23 (2.2)	0.57 (0.52, 0.62)					
US Norms	23 (2.2)	131 (12.6)	735 (70.8)	129 (12.4)	20 (1.9)						

* n refers to the number of tests administered, not individual children.

** Confidence interval.