

# Methods of Neurodevelopmental Assessment in Children with Neurodegenerative Disease: Sanfilippo Syndrome

Kathleen A. Delaney · Kyle R. Rudser ·  
Brianna D. Yund · Chester B. Whitley ·  
Patrick A.J. Haslett · Elsa G. Shapiro

Received: 25 January 2013 / Revised: 07 August 2013 / Accepted: 25 September 2013 / Published online: 5 November 2013  
© SSIEM and Springer-Verlag Berlin Heidelberg 2013

**Abstract** Objectives: (1) Develop a methodology for obtaining reliable cognitive and developmental data in children with neurodegenerative disease and cognitive impairment and in turn monitor disease state and treatment outcomes. (2) Demonstrate validity of age-equivalent scores.

Methods: We present guidelines for obtaining accurate test scores in low-functioning and behaviorally disruptive pediatric patients, followed by a method validation study: (1) using disease-specific protocols to assess salient aspects of the known phenotype, (2) selecting appropriate tests, (3) managing behavior, and (4) using age-equivalent scores on standardized tools. We used the Bayley Scales of Infant Development-III or Kaufman Assessment Battery for Children-II with a group of 25 children with mucopolysaccharidosis type IIIA (MPS IIIA or Sanfilippo syndrome type A) with dementia. To demonstrate concurrent validity, we used the Vineland Adaptive Behavior Scales-II, comparing parent-reported age-equivalent scores (AEs) with those of the cognitive measures.

Results: We were successful in obtaining cognitive age-equivalents for 25 patients with MPS IIIA including those with severe behavioral disruption and a correlation of 0.95 was obtained comparing scores on the parent measure with cognitive age-equivalents validating the age-equivalent approach.

Conclusion: An approach to the assessment of severely impaired children including those with behavioral disruption was implemented and is applicable to children with other severe neurological diseases. This approach will enhance the assessment of disease progression and monitoring of treatment outcome in clinical trials.

## Background

Obtaining reliable and accurate neurocognitive assessment data is a challenge in children with neurodegenerative disorders who have dementia or are very low functioning, have disruptive, noncooperative behavior, or have physical/sensory disabilities. These challenges and the limitations of available measures have been documented previously in patients with lysosomal diseases (Martin et al. 2008). Solutions that will result in precise assessment are essential to assess disease progression and the effects of treatment. We present here an example of how such data can be obtained using a pragmatic, disease-specific approach using age-equivalent scores. Our ultimate goal is to obtain precise data about measurement of disease progression and ultimately treatment outcomes. This becomes a more urgent problem as new treatments emerge for rare neurogenetic diseases.

Standard psychology practice currently does not support the use of age-equivalent scores (AgeEqSs) as their statistical properties are inadequate and can be misleading (Naglieri and Goldstein 2009). AgeEqSs do not consider

---

Communicated by: Frits Wijburg, MD, PhD

---

Competing interests: None declared

---

K.A. Delaney · B.D. Yund · C.B. Whitley · E.G. Shapiro  
Department of Pediatrics, University of Minnesota, Minneapolis, MN,  
USA

K.R. Rudser  
Department of Biostatistics, University of Minnesota, Minneapolis,  
MN, USA

P.A.J. Haslett  
Shire, Lexington, MA, USA

K.A. Delaney (✉)  
MMC 486 Delaware St., SE Minneapolis, MN 55455, USA  
e-mail: delan011@umn.edu

the range of normality in contrast to standard scores which establish a range of normal performance (Maloney and Larrivee 2007). Standard deviations vary from age to age and from test to test. The scale units of AgeEqSs are ordinal and the intervals between units are unequal (e.g., an increase of 6 months at one age will differ from an increase at another age) (Lawrence 1992). For these reasons, psychologists have resisted the use of AgeEqSs and have primarily used age-stratified normative data to obtain standard scores. However, those caveats may hold for higher functioning children; for younger more severely impaired children having an ordinal measure that tracks change over time is crucial. Developmental growth curves can be constructed from AgeEqSs to track changes; pediatricians commonly use such curves to monitor height, weight, and head circumference. The difference and challenge here is that, unlike measuring anthropometric measurements (e.g., height, weight, and OFC), uneven skill development within children with neurodegenerative disease challenges the ability to measure their overall development level especially using standard scores.

There are several reasons that standard scores should not be used in very low-functioning children. The standard scores on most tests have a floor; most low-functioning children fall below this floor which makes them insensitive to any change. Furthermore, for a child with dementia, even if a standard score can be obtained at an initial visit, it is very possible that at the next evaluation the child will no longer fall within the range of standard scores available for that measure.

In contrast, AgeEqSs are easily interpretable and allow for precise longitudinal monitoring of illnesses in children with cognitive decline and/or who are very low functioning. Change in rate of growth or decline can be monitored with respect to the disease process or the effects of treatment. One can determine whether a child continues to gain milestones, has reached a plateau, or is losing skills. However, AgeEqSs do not necessarily imply that the child has behavior and development typical of that age, e.g., a child with Sanfilippo syndrome who has an AgeEqSs of 12 months does not necessarily present as a chronologically older child who acts like a 12-month-old child.

In a new development in psychology, the most recent edition of the Bayley Scales of Infant and Toddler Development-III (BSID-III) has added such developmental growth curves to the manual, thus allowing one to track AgeEqSs over time (Bayley 2006). The use of AgeEqSs to assess the development of children with dementia was initially proposed by Shapiro and Klein in 1993 (Shapiro and Klein 1993) and has been endorsed in other discussions of neurodegenerative diseases (Shapiro and Balthazor 1999; Ziegler et al. 2010). As early as 1985, Volkmar et al. recommended the use of age-equivalent scores as a precise measure of function in low-functioning individuals allow-

ing comparisons across domains of function (Volkmar et al. 1987). AgeEqSs have been used in neurodegenerative diseases, primarily the mucopolysaccharidoses, to examine treatment effects using both cognitive (Staba et al. 2004; Peters et al. 1996, 1998; Wraith et al. 2007) and Vineland Adaptive Behavior Scales (Sparrow et al. 2005) age-equivalents. While the use of cognitive AgeEqSs for the BSID II, the Mullen Scales of Early Learning (Mullen 1995), or the Griffiths Mental Development Scales (Huntley 1996) has been common in very low-functioning children, the problems and challenges surrounding cognitive assessment in childhood dementia have not been carefully addressed nor have methods been delineated for obtaining and verifying this information. The validity and sensitivity of AgeEqSs needs to be demonstrated to use them in treatment trials.

We developed a neurodevelopmental assessment approach with applicability to any neurodegenerative disease based on our experience with mucopolysaccharidosis type IIIA (MPS IIIA) or Sanfilippo syndrome. MPS IIIA is associated with dementia and severe behavioral disruption. It is a rare (about 1 in 100,000 births) (Meikle et al. 1999; Poorthuis et al. 1999; Baehner et al. 2005) autosomal recessive lysosomal storage disease caused by absence of enzyme heparan-N-sulfatase (sulfamidase) necessary for degradation of the glycosaminoglycan (GAG) heparan sulfate in lysosomes which then accumulates in the lysosomes of neurons and glial cells. MPS IIIA is a neurodegenerative disease with a variable trajectory in neurocognitive decline with respect to age and time of diagnosis; however, this decline and the variability in decline are not well documented. Deficits in language development, motor skills, and intellectual development have been reported as early as 2 years of age (Cleary and Wraith 1993). Yet precise measurement of the decline of intellectual skills is absent, likely due to the difficulties in testing these children given their tendency toward behavioral noncompliance (Valstar et al. 2011). Abnormal behaviors have been described as aggressive and hyperactive with sleep disturbance (Cleary and Wraith 1993; Valstar et al. 2011; Meyer et al. 2007; Fraser et al. 2002). Autistic-like symptoms and loss of normal fear have also been described (Heron et al. 2011). In a recent study, (Heron et al. 2011) found the average lifespan for MPS IIIA to be 15.4 years (Heron et al. 2011).

In the classic form of MPS IIIA, symptoms appear to arise between 2 to 6 years of age although diagnosis often lags behind the earliest symptoms (Cleary and Wraith 1993). Some MPS IIIA patients, who have clinical onset and diagnosis after age 6, have been described by Hopwood (2007) and by (Heron et al. 2011) with slower decline and an unknown disease progression. Instruments used for assessment need to encompass a wide range of abilities and ages as previously established by Valstar et al. (2011).

## Guidelines

Neurocognitive assessment has been a challenge for investigators working with MPS IIIA due to the patients' tendency toward behavioral disruption, their noncompliance, and the lack of disease familiarity of the examiner. Additional challenges include the unavailability of tests/methods sensitive to change in very low-functioning children (Valstar et al. 2011). We sought to develop an approach that would deal with these obstacles using the following four guidelines:

### Guideline 1. Protocols and Testing Approaches Must Be Disease Specific

Understanding the disease phenotype is crucial to sensitive testing. We began with clinical behavioral observations of children with MPS III to determine why testing was difficult, and with a goal of finding tasks that they could perform. Disease-specific neurological and medical impediments to testing were defined. Hearing problems were almost universal which affects language-based test results. The use of hearing aids was important in obtaining reliable data. Furthermore, increasing auditory agnosia (inability to comprehend word meaning) is evident as the disease progresses. The child previously understood the word and used the word, but no longer is able to understand its meaning even while continuing to say the word. For these reasons, language-based tests as a measure of cognition were avoided and nonverbal assessment was necessary.

Motor apraxia increases as the neurologic disease progresses. While children with MPS IIIA can spontaneously perform motor activities, they are unable to perform these activities to either verbal instruction or imitation. Consideration of this factor is important as many nonverbal tests require motor output.

Behavioral noncompliance often occurs when the child is unable to do a task or is frustrated. Testing needs to be easy enough for success. Behavioral noncompliance is not the cause of inability to perform, but rather the inability to perform leads to behavioral disruptions and noncompliance which may be typical of the way the MPS IIIA child responds to frustration.

**MPS IIIA Testing Challenge:** Psychologists with no experience with MPS IIIA assume behavior noncompliance is the cause of the child's poor performance. Parents report the child is able to perform the item at home, but cannot perform it in the clinic; the child seems to lack cooperation on items that appear easy.

**Solution:** The neurologic conditions of motor apraxia and auditory agnosia in MPS IIIA cause lower scores than the tester might expect when observing spontaneous

behavior. Children cannot perform using instruction or imitation. Also, other medical conditions in MPS IIIA can interfere with test performance. Hearing aids and glasses must be brought to the clinic and used. A recent audiology consult to assess hearing status should precede testing. If the child is not well or has a cold, testing should be delayed if possible. Use of pacifiers or "chewies" should be allowed during testing if the child uses them at home. Allow extra time. Children with neurodegenerative disease often have a slower response to requests and it is critical to allow for more time on a task before determining their success. At the same time, spending too much time on any one task or item could lead to frustration during the evaluation.

Do not allow the visit to occur until at least 48 h after sedation or anesthesia or after a recent seizure. Consider whether current medications affect cognition.

### Guideline 2. Criteria for Test Selection Includes Appropriate Difficulty Level, Disease-Relevant Domains of Content, and Familiarity with the Test by Psychologists

Our goal is to include tests of cognition, language, motor ability, memory, attention, adaptive function, and behavior. In choosing tests for a protocol, keep the number of measures of any domain to a minimum to increase the validity and comparability of results. In MPS IIIA, the range of cognitive impairment may vary from severe to near-normal in very young children with an age range from infancy to young adulthood. No single measure meeting our criteria covered that range so we needed to use two different measures. Where feasible this should be avoided. No test has standardized scores for very low levels; thus, age-equivalent scores were needed that are psychometrically sound to make the tests comparable.

Because sufficient sample size cannot be accrued in one center if the disease is rare, a protocol should be feasible in multiple centers and in multiple countries. Decreased accuracy and reliability of data in multicenter studies must be offset with precise and simple measurements.

We reviewed many cognitive tests and rejected them. The Mullen Scales of Early Learning (MSEL) (Mullen 1995) is usually our preferred measure because it is comprised of a variety of domains, and each scale begins with birth and is validated up to 7 years of age. However, while used widely in the USA, the MSEL is not familiar to international researchers and is not translated into other languages. The Griffiths Scale of Mental Abilities is unknown in the USA even though it is widely used in the UK and other English-speaking countries (also translated

into Canadian French and German) (Huntley 1996). Another test we considered was the Leiter Scales (Roid and Miller 1997). The age-equivalent scores were not well standardized and it was somewhat outdated. The Stanford Binet Intelligence Scale included too many verbal-based subtests and did not go below an age-equivalent of 2 years and is only available in English (Roid 2003) and the Wechsler Preschool and Primary Scale of Intelligence was too difficult with a high floor, although available in many languages (Wechsler 2002). Most standard IQ tests have a “floor” of a standard score of 40 or 50 (4 or 3.3 standard deviations below the mean). The Differential Abilities Scale-II (DAS-II) (Elliott 2007) is both difficult to administer and to score, and does not have an international presence (although a Spanish translation exists for the preschool level). Also, the DAS II subtests change from one age grouping to another making it difficult to create a developmental growth trajectory for children who straddle those age groups. The Kaufman Assessment Battery for Children – Second Edition is widely used in cross-cultural research because much of the test is nonverbal and it has been translated into more than 15 languages. Although the test is based on psychological theory which may not apply to such severe impairment, it has a nonverbal scale that is widely used with hearing and language impaired children (Kaufman and Kaufman 2004).

**MPS IIIA Challenge:** Find tests measuring cognitive ability, adaptive function, and other domains that have an appropriate range of difficulty for MPS IIIA which are widely used with age-equivalent scores that are psychometrically sound.

**Solution:** Two cognitive tests were chosen for MPS IIIA: Bayley Scales of Infant Development – Third Edition (BSID-III) (Lawrence 1992) and Kaufman Assessment Battery for Children – Second Edition (KABC-II) (Elliott 2007) and one adaptive scale, The Vineland Adaptive Behavior Scales – Second Edition (VABS-II) (Wraith et al. 2007) for the following reasons: (1) universality of use, (2) availability of age-equivalent scores for severely impaired children, (3) nonverbal content as on the cognitive scale on the BSID and the nonverbal scale of the KABC-II for higher functioning children, and (4) availability of supplementary language and motor assessment (both domains on the BSID-III and some language on the KABC-II). The VABS-II, a parent-reported outcome, was selected as an adaptive measure for the same reasons.

Here is the algorithm we developed to choose which measure to administer for each MPS IIIA patient: The VABS-II is always given first.

- If chronological age is under 42 months, the BSID-III was selected.
- If chronological age > 42 months and age-equivalent score on the VABS-II > 42 months, the KABC-II is selected. Forty-two months was chosen so that it would allow considerable decline to occur so that, over the study interval of one year, the child would not fall below the base of the test (lowest possible score).
- If chronological age > 42 months and VABS-II age-equivalent is between 36 and 42 months, start with the Triangles subtest on the KABC-II; if the child cannot do a total of two items on the Triangles subtest, we try one other nonverbal subtest. If the child cannot successfully perform three items, we fall back to the BSID-III. We always start with item #1 on every KABC-II subtest for this group.
- If chronological age is > 42 months and mental age-equivalent is < 36 months on the VABS-II, the BSID-III is administered with the cognitive age-equivalent score as the primary outcome.

### Guideline 3. Testers Must Be Able to Deal Effectively with Behavioral Difficulties and To Understand the Child’s Behavior in the Context of the Disease Process

Nothing is better than experience in testing children with very impaired and behaviorally dysregulated behavior. In children with dementia, often commonly utilized behavioral paradigms such as reinforcement contingencies do not work. Many factors contribute to behavioral noncompliance and should be considered such as diurnal variation in behaviors, lack of sleep and fatigue (often from travel to the testing facility), previous medical evaluations, lack of hearing aids and glasses, lack of ability to perform the task for reasons related to the disease, and frustration.

Here is a short list of testing pointers: avoiding much or too complicated verbalization when administering tests, being sensitive to the physical needs of the child (letting the child have a pacifier or chewie), letting the child up out of the chair, letting the child engage in repetitive behaviors (while observing and recording them) but then distracting the child away from these behaviors, and finding behaviors to reward. Having the parent in the testing room may contribute to behavioral disruption or may help prevent it; a parent’s presence is a necessity in about a half of children with MPS III.

**MPS IIIA Challenge:** Accomplishing neurodevelopmental testing in the face of behavioral dysregulation and lack of compliant behavior.

**Solution:** Our tests apply the principles mentioned before and where possible:

- Test the child at the optimum time, (i.e., mainly in the morning)
- Give the child ample time to respond to items in a child-friendly environment (defined as nonmedical, appropriate furniture, windowless, small and a quiet, relatively stimulus-free room).
- Consult with parents prior to testing about rewards, etc., that may be useful. Contingencies are difficult for these children to understand; you must “catch” the right behavior to reward it.
- Parents are instructed on what to do in the testing room if they observe during the session. Parents can be a big help but, if they interfere, should be told firmly not to do so.
- Make sure that easy items are given first and items increase in difficulty gradually. Always assume that noncooperation reflects frustration from items that are too difficult. Start with something the child can do. This provides an opportunity for the child to become acclimated to testing and also puts the parent at ease. Allow the child to first play with simple and engaging test materials without risking test integrity. If the child is fond of a particular toy and becomes restless, use that toy (assuming it is not to be used later in the testing) as a distracter between tasks or measures.
- Assess and implement strategies based on the child’s temperament and approach to testing. For example, if the child is easily overstimulated, e.g., startles easily, cries a lot, avoids eye contact, then use a quiet tone, slower movements, and present one item at a time. If the child is disengaged, then use exaggerated movements, encouragement, and praise to help elicit a response.
- If the child is particularly disturbed by or disinterested in the object, take it away and attempt that item later.

#### **Guideline 4. Tests Must Be Scorable for all Participants Using Age-Equivalents**

The BSID-III, KABC-II, and VABS-II have a lower limit for which the test was standardized. For the BSID-III and KABC-II, the lowest standardized score is 40 and for the VABS-II it is 20. Many of the children in this study are considerably below that limit. Consequently we chose to use age-equivalent scores (AgeEqSs) for the reasons

outlined in the introduction. Using AgeEqSs we can examine growth rates for treated and untreated children across tests. Of course, there are some caveats to cross-test comparison; the nature of the normative samples and the years since the norming took place may vary across tests.

We can also obtain a developmental quotient (DQ) using AgeEqSs in the numerator and chronological age at testing as the denominator multiplied by 100. This was the method used when IQ tests were first developed, but these ratio IQs were abandoned because the standard deviations vary by age. In modern psychological tests, a standard score is computed with a set standard deviation of 15 and a mean of 100. DQs are useful when an approximation to an IQ is needed demonstrating the gap between the child with a neurodegenerative condition and a typically developing child of the same age. However, AgeEqSs provide more detailed information about changes in rate of development due to disease or treatment.

**MPS IIIA Challenge:** As no standard score data are available for very low-functioning individuals, we have used AgeEqSs. Any nonstandard approach to scoring neuropsychological tests must be validated.

**Solution:** We sought to validate the age-equivalent approach to demonstrate that our approach is reliable and yields accurate data. The VABS-II, a parent report measure, yields AgeEqSs making it possible to compare it to scores on a directly administered measure (BSID-III and KABC-II), thus obtaining a measure of concurrent validity. Our results are indicated in the study outlined below:

## **Methods**

### **Patients**

Twenty-five patients with documented Sanfilippo syndrome type A were recruited to this single-center study supported by Shire (A 12-month Longitudinal, Prospective, Natural History Study of Patients with Sanfilippo syndrome type A (MPS IIIA), Clinicaltrials.gov reference number, NCT01047306).

Patients were recruited from our own clinics and through patient advocacy groups and Clinicaltrials.gov. Inclusion criteria were as follows: (1) documented deficiency in HS enzyme activity of less than or equal to 10 % of the lower limit of the normal range as measured in fibroblasts or leukocytes and normal enzyme activity level of at least one other sulfatase (to rule out multiple sulfatase deficiency) as measured in fibroblasts or leukocytes; (2) a developmental age greater than or equal to 1 year of age on the interview

**Table 1** Descriptive data

Measure $N = 25$ for all measures	Mean (standard deviation)
Age: mean (S.D.)	75.0 months (53.4) range 13–220 months
Cognitive age-equivalent (BSID or KABC)	23.0 months (12.9) range 7–58 months
Developmental quotient (mean age-equivalent of BSID or KABC/chronological age)	44.6 (28.4) range 3–91
Vineland age-equivalent (mean of subscales except motor domain)	23.8 months (12.8) range 6–61 months
Vineland developmental quotient (mean of age-equivalents except motor domain/chronological age)	44.4 (25.3) range 3–95
Standard score using normative data on the Vineland	63.0 (16.8) range 25–95

form of the VABS-II (Sparrow et al. 2005); and (3) the patient was determined to be medically stable to accommodate the protocol requirements. Relevant exclusion criteria included non-MPS IIIA-related central nervous system impairment or behavioral disturbance, blindness, deafness, poorly controlled seizures, and other treatments (e.g., hematopoietic cell transplant, investigational drugs or devices). This protocol was approved by the Institutional Review Board and all parents of patients signed consents.

#### Materials and Procedure

Twenty-two patients were administered the Bayley Scales of Infant Development-III (BSID-III) and three had the Kaufman Assessment Battery for Children-II (KABC-II). Decisions were made according to the rules specified above. While the BSID-III has motor and language scales and they were also administered, only scores for the cognitive scale are reported here. The cognitive scale of the BSID-III has a split-half reliability coefficient of 0.91 (Bayley 2006). The KABC II has many subtests and scales. We administered only the Nonverbal Index which includes five different subtests: Story Completion, Triangles, Block Counting, Pattern Reasoning, and Hand Movements. The internal consistency as a measure of reliability of the Nonverbal Index is 0.92 (Elliott 2007).

Parents of all 25 patients were administered the Vineland Adaptive Behavior Scale – Second Edition-Interview format (VABS-II). These data were used to determine eligibility as well as to assess adaptive function.

**Procedure:** Two independent examiners interviewed parents to administer the VABS-II. The two examiners carried out 16 and 9 VABS-II interviews respectively prior to cognitive testing. Importantly, both examiners who administered the cognitive tests were experienced with MPS and other neurodegenerative diseases. The primary examiner (KD) tested 22/25 patients. She supervised the other examiner who assessed patients who had fewer behavioral difficulties or were older. All scoring was vetted by KD. All testing was carried out at the Center for Neurobehavioral Development, University of Minnesota

and was done in the morning of the second day of the patient's visit (to ensure recovery from travel).

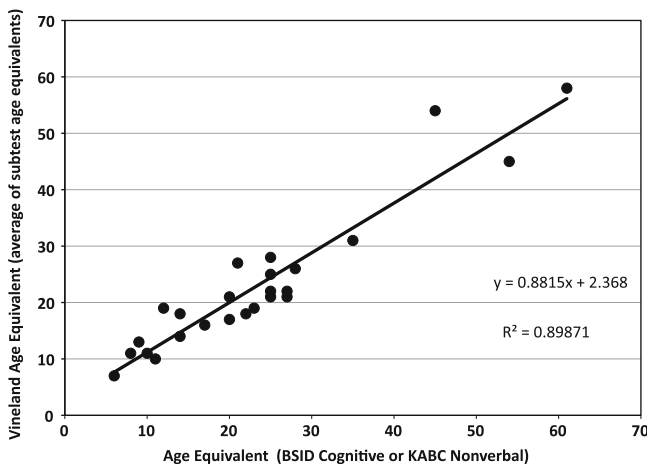
**Scoring and statistical methods:** Either the BSID-III cognitive age-equivalent or the KABC-II nonverbal age-equivalent was calculated for each patient. The mean age-equivalent of the subtests of the Nonverbal Index was calculated as the best measure of language-free cognition on the KABC-II in the older and less impaired children which was most equivalent to the BSID cognitive age-equivalent.

For the VABS-II, we calculated the mean age-equivalent over the subtests, excluding the VABS motor scale, to generate a score as equivalent as possible to those of the cognitive tests. The goal was to establish concurrent validity between the VABS-II and the cognitive tests. Concurrent validity is shown when scores on a test are highly associated with scores on another measure which tests a similar set of skills or abilities. Both tests presumably reflect related constructs. We calculated a correlation coefficient between the cognitive age-equivalent (either BSID-III cognitive or KABC-II nonverbal) and the VABS-II age-equivalent.

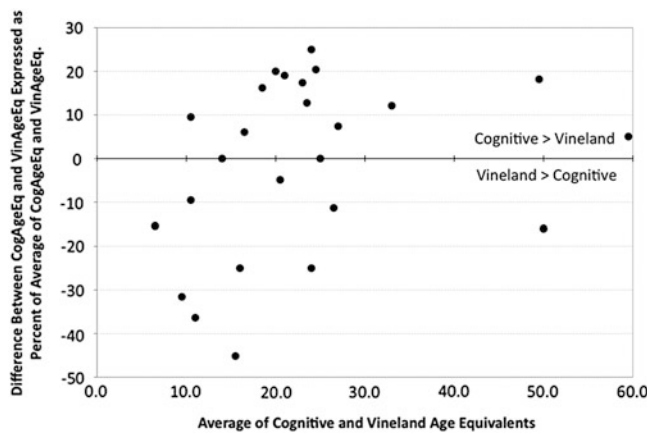
However, a high correlation does not necessarily imply that both measures will show agreement in level of scores between the two methods of assessment. A Bland-Altman (Bland and Altman 1986) style plot was used to examine the agreement between these two different tests with the mean of the two assessments on the x-axis and the difference as a percent of the mean on the y-axis.

#### Results

Both cognitive ability and the VABS-II age-equivalent scores (AgeEqSs) yielded similar mean scores and standard deviations (See Table 1). Cognitive ability as measured by the DQ (ratio of cognitive age-equivalent to chronological age) was discrepant from the Adaptive Behavior Composite standard score but not the DQ (mean of the age-equivalents for all subtests divided by chronological age) on the VABS-II (Table 1).



**Fig. 1** Association of BSID or KABC age-equivalent score with Vineland age-equivalent score



**Fig. 2** Percent difference in VABS and cognitive scores by cognitive age-equivalent (Bland-Altman graph)

A correlation of 0.95 was found between AgeEqSs on the cognitive tests and on the VABS-II. See Fig. 1.

From the Bland-Altman graph, we conclude that the parent ratings of the lower functioning children, when compared to the mean of both measures, have a tendency toward more positive scores on the VABS-II. See Fig. 2.

**Discussion**

The two aims of this study were as follows: (1) to provide guidelines for the clinical collection of neurodevelopmental data in children with severe neurodegenerative diseases and (2) to demonstrate that the accuracy of such data even with the most severely impaired children can be precise, thus making it useful for understanding disease progression as well as for sensitive treatment monitoring. We found that following such guidelines we could measure disease

progression with accuracy and precision even in children who are severely impaired.

We have demonstrated the concurrent validity of AgeEqSs. In our results, the positive correlation of the AgeEqSs on the BSID-III/KABC-II with the VABS-II likely is due to several factors. The first is the extensive experience in our clinics with cognitively normal children, and children with lysosomal and other dementing illnesses. Our transplant program and our lysosomal disease program have introduced us to hundreds of children with these diseases. Our tester KD is extraordinarily skilled at getting cooperation and obtaining accurate scores on these tests. It is highly unlikely that a tester never having seen a child like this will be able to achieve that breadth of understanding and accuracy of assessment. The second is the very controlled circumstance of our testing procedures: time of day, testing environment, and organization of the testing experience. Also, we note that in more than half the cases, parent report on the VABS-II was obtained by someone other than the tester with no knowledge of the test results, making those scores in part independent of each other. We conclude that accurate developmental data that is sensitive to change can be obtained, but that adherence to the above guidelines can increase accuracy.

In 1992, Raggio, Massingale, and Bass (Raggio et al. 1994) administered the VABS and the BSID (both first editions of those tests) to a sample of high-risk infants who were suspected of developmental delay to answer the question of whether standard scores or age-equivalents were a better measure. The standard scores obtained on their sample were higher than the age-equivalents and the MDI, and the authors concluded, as did we, that AgeEqSs are more accurate. These authors propose that age-equivalents are better measures for severely impaired children because of the imprecision of very low standardized scores on the VABS. Our results support that finding and extend it to a larger age range and up-to-date tests.

We found a tendency toward more positive scores on the VABS-II. This discrepancy is more pronounced in the most impaired children with the lowest cognitive age-equivalents. Thus, even while using age-equivalents instead of standard scores, getting precise scores may be difficult especially on the VABS-II. Although these results support the use of direct measurement of cognitive function, the easily administered VABS-II could be used to estimate cognitive ability in settings where the BSID-III might be difficult, for lack of availability or expertise.

A specific explanation for the difference in performance between the parent report and direct measurement of the child’s performance in low-functioning children could be the motor apraxia and auditory agnosia that were observed in MPS IIIA. These symptoms may prevent the child from performing an item to instruction or imitation, while

spontaneously they may be able to produce this behavior in the home environment; we have reported this elsewhere (Delaney et al. 2013). Although we cannot determine which score is more “correct,” a score in a standardized testing situation such as the BSID-III or KABC-II reflects the consequences of the disease and should change over time with disease progression.

Each test assesses different aspects of cognitive development. The lack of overlap of what test items measure is a limitation of tracking from one test to another using AgeEqSs. Tests of infant and toddler development assess primarily visual and motor development; tests that start at age 2 or 3 will have more problem-solving and reasoning tasks (Sigelman et al. 2011). This may result in a discontinuity in what the AgeEqSs are measuring. However, the association with a parent report suggests that even though the items are quite different, a similar developmental trajectory can be identified.

Several papers have been published in the past regarding cognitive methods of assessment in lysosomal storage diseases (Martin et al. 2008; Shapiro et al. 1995). Valstar et al. addressed the issue of methods to assure reliability using familiar environments to decrease behavioral problems as well as tests for different levels of development (Valstar et al. 2011). Previously AgeEqSs have been used in some studies of cognitive and adaptive development (Staba et al. 2004; Peters et al. 1996, 1998; Wraith et al. 2007; Valstar et al. 2011; Bjoraker et al. 2006) despite their lack of acceptability in the psychology literature. What is new here? This study has (1) demonstrated the concurrent validity of AgeEqSs in cognitive ability testing, and (2) provided detailed specific clinical guidelines for testing behaviorally and cognitively impaired children. This is important as precise measurement of functional outcomes will be necessary as treatments for these children are developed.

The approach to testing we have taken with MPS IIIA is also applicable to multicenter studies of children with dementia or severe cognitive impairment in other conditions such as in late infantile Batten disease, Tay-Sachs disease, Krabbe disease, and Niemann-Pick disease type C. An in-depth understanding of how the disease alters the child's limitations should be the basis of test selection. Specific challenges that may arise due to neurological or medical concomitants of these disorders should be the foundation of the testing approach.

New treatments for genetic illnesses of children are now in trials or may soon be available. Enzyme replacement, gene therapy, stem cell therapies, small molecule medications, anti-inflammatories, and chaperone therapies will all require assessment of change over time. This is challenging because many of these children plateau in their development for long periods of time with sudden declines followed by another plateau. It requires natural history

studies that trace the developmental growth curves of these patients without treatment to determine whether a treatment might alter the rate of growth or the slope of development. We propose that using guidelines for acquisition of neurodevelopmental data in severely impaired children to obtain AgeEqSs for test scores yields a useful, easily interpretable approach to assessing developmental trajectories and longitudinal change.

**Acknowledgements** The authors would like to acknowledge the support of Shire Pharmaceuticals and the Lysosomal Disease Network.

The Lysosomal Disease Network (U54NS065768) is a part of the National Institutes of Health (NIH) Rare Diseases Clinical Research Network (RDCRN), supported through collaboration between the NIH Office of Rare Diseases Research (ORDR) at the National Center for Advancing Translational Science (NCATS), the National Institute of Neurological Disorders and Stroke (NINDS) and National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Summary Statement

A validated method of neurodevelopmental assessment in children with Sanfilippo and other neurodegenerative diseases.

## References

- Baehner F, Schmiedeskamp C, Krummenauer F et al (2005) Cumulative incidence rates of the mucopolysaccharidoses in Germany. *J Inherit Metab Dis* 28:1011–1017
- Bayley N (2006) Bayley scales of infant and toddler development—third edition. Psychological Corporation, San Antonio
- Bjoraker K, Delaney K, Peters C et al (2006) Long term outcomes of adaptive functions for children with MPS I. *J Beh Dev Peds* 27:290–296
- Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 327:307–310
- Cleary MA, Wraith JE (1993) Management of mucopolysaccharidosis type III. *Arch Dis Child* 69:403–406
- Delaney K, Nestrail I, Yund B et al (2013) Motor function decline and motor apraxia in Sanfilippo syndrome. *Mol Genet Metab* 108:S34
- Elliott CD (2007) Differential ability scales—second edition. Psychological Corporation, San Antonio
- Fraser J, Wraith JE, Delatycki MB (2002) Sleep disturbance in mucopolysaccharidosis type III (Sanfilippo syndrome): a survey of managing clinicians. *Clin Genet* 62:418–421
- Heron B, Mikaeloff Y, Froissart R, et al (2011) Incidence and natural history of mucopolysaccharidosis type III in France and comparison with United Kingdom and Greece. *Am J Med Genet A* 155A:58–68
- Hopwood J (2007) Sanfilippo syndrome: clinical genetic diagnosis and therapies. In: Barranger J (ed) *Lysosomal storage disorders*, Ch. 26. Springer, New York, pp 415–432
- Huntley M (1996) *The Griffiths mental development scales: from birth to 2 years*, 2nd ed. The Test Agency Limited, Oxford



- Kaufman AS, Kaufman NL (2004) Manual for the Kaufman assessment battery for children—second edition (KABC-II) comprehensive form. American Guidance Service, Circle Pines
- Lawrence CW (1992) Assessing the use of age-equivalent scores in clinical management. *Lang Speech Hear Serv Sch* 23(1):6–8
- Maloney ES, Larrivee LS (2007) Limitations of age-equivalent scores in reporting the results of norm-referenced tests. *Contemporary Issues Commun Sci Disorder* 34:86–93
- Martin HR, Poe MD, Reinhartsen D, Pretzel RE, Roush J, Rosenberg A, Dusing SC, Escolar ML (2008) Methods for assessing neurodevelopment in lysosomal storage diseases and related disorders: a multidisciplinary perspective. *Acta Paediat* 97:69–75
- Meikle PJ, Hopwood JJ, Clague AE, Carey WF (1999) Prevalence of lysosomal storage disorders. *JAMA* 281:249–254
- Meyer A, Kossow K, Gal A et al (2007) Scoring evaluation of the natural course of mucopolysaccharidosis type IIIA (Sanfilippo syndrome type A). *Pediatrics* 120:e1255–e1261
- Mullen EM (1995) Mullen scales of early learning. American Guidance Service, Circle Pines
- Naglieri J, Goldstein S (2009) Practitioner's to guide assessing intelligence and achievement. Wiley, Hoboken
- Peters C, Balthazor M, Shapiro E et al (1996) Outcome of unrelated donor bone marrow transplantation in forty children with Hurler syndrome. *Blood* 87:4894–4902
- Peters C, Shapiro E, Anderson J et al (1998) Hurler Syndrome: II. Outcome of HLA-genotypically identical sibling and HLA-haploidentical related donor bone marrow transplantation in fifty-four children. *Blood* 91:2601–2608
- Poorthuis BJ, Wevers RA, Kleijer WJ et al (1999) The frequency of lysosomal storage diseases in The Netherlands. *Hum Genet* 105:151–156
- Raggio DJ, Massingale TW, Bass JD (1994) Comparison of Vineland adaptive behavior scales- survey form age equivalent and standard score with the Bayley Mental Development Index. *Percept Mot Skills* 79:203–206
- Roid GH (2003) Stanford-Binet intelligence scales (5th ed). Riverside, Itasca, NY
- Roid GH, Miller LJ (1997) Leiter international performance scale-revised: examiner's manual. Stoelting Co., Wood Dale
- Shapiro E, Balthazor M. Metabolic (1999) and neurodegenerative disorders of childhood. In: G Taylor, D Ris, K Yeates (eds) *Pediatric neuropsychology: research, theory and practice*. Guilford Press, New York. pp 171–205
- Shapiro E, Klein K (1993) Childhood dementia: neuropsychological assessment and treatment of degenerative childhood diseases. In: Tramontana MG, Hooper SR (eds) *Advances in child neuropsychology, Volume III*. Ch. 4. Springer, New York, pp 119–171
- Shapiro EG, Lockman LA, Balthazor M, Krivit W (1995) Neuropsychological outcomes of several storage diseases with and without bone marrow transplantation. *J Inherit Metab Dis* 18:413–429
- Sigelman CK, Rider EA (2011) *Life-span human development – sixth edition*. Wadsworth Cengage Learning, Belmont
- Sparrow SS, Cicchetti DV, Balla DA (2005) *Vineland adaptive behavior scales, 2nd edn*. Psychological Corporation, San Antonio
- Staba SL, Escolar ML, Poe M et al (2004) Cord blood transplants from unrelated donors in patients with Hurler's syndrome. *NEJM* 350:1060–1069
- Valstar MJ, Marchal JP, Grootenhuys M et al (2011) Cognitive development in patients with mucopolysaccharidosis type III (Sanfilippo syndrome). *Orphanet J Rare Dis* 6:43
- Volkmar FR, Sparrow SS, Goudreau D, Cicchetti DV, Paul R, Cohen DJ (1987) Social deficits in autism: an operational approach using the Vineland Adaptive Behavior Scales. *J Amer Acad Child Psychiat* 26:156–165
- Wechsler D (2002) *Wechsler preschool and primary scale of intelligence, 3rd edn*. Psychological Corporation, San Antonio
- Wraith JE, Beck M, Lane R et al (2007) Enzyme replacement therapy for mucopolysaccharidosis I patients less than 5 years old: Results of a multinational study of recombinant human alpha-L-iduronidase (laronidase). *Pediatrics* 120:e37–e46
- Ziegler R, Shapiro E (2010) Metabolic neurodegenerative diseases across the lifespan. In: Donders J, Hunter S (eds) *Principles and practice of lifespan developmental neuropsychology*, ch. 15. Cambridge University Press, Cambridge, pp 427–448