CONCURRENT VALIDITY AND RESPONSIVENESS OF THE PEABODY DEVELOPMENTAL MOTOR SCALES - 2 IN INFANTS AND CHILDREN WITH POMPE DISEASE UNDERGOING ENZYME REPLACEMENT THERAPY

Dawn Phillips PT, PhD

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree Doctor of Philosophy in the School of Medicine, Program in Human Movement Science.

Chapel Hill
2012

Approved by:

Angela. E. Rosenberg, PT, Dr PH

Barry Byrne, MD, PhD

Vicki Mercer, PT, PhD

Rebecca E. Pretzel, PhD

Todd Schwartz, Dr PH
ABSTRACT

DAWN PHILLIPS: Concurrent Validity and Responsiveness of the Peabody Developmental Motor Scales - 2(PDMS-2) in Infants and Children with Pompe Disease undergoing Enzyme Replacement Therapy (Under the direction of Angela E. Rosenberg, PT, Dr PH)

Purpose: To examine the responsiveness of the PDMS-2 in children diagnosed with Pompe disease who had different levels of functional mobility. To examine the concurrent validity between the PDMS-2 and the Alberta Infant Motor Scale (AIMS), the Pediatric Evaluation of Disability Inventory (PEDI) and the Pompe PEDI (PPEDI) in children diagnosed with Pompe disease. Methods: A secondary analysis was completed of the Genzyme Corporation Pompe efficacy trials for Myozyme. The children were divided into two functional groups, independent ambulators (group 1), and children who required use of assistive devices to ambulate or were unable to ambulate (group 2). Results: A significant difference was present between the mean PDMS-2 subtest and Gross Motor Composite percentage scores at baseline and at week 52 when the whole group was combined. Responsiveness measured by the factor of Time and Function revealed Locomotion subtest statistically significant mean percentage change within both functional groups from baseline to week 52 and between groups at both time periods. No significant within group or between group differences were seen for the Stationary subtest. The Object Manipulation
subtest demonstrated significant change within group one from baseline to week 52 and between the functional presentations at week 52. The age equivalent correlations between the PDMS-2 and AIMS indicate a good to excellent relationship for all subtests except the Object Manipulation subtest. A four-month range in age equivalent scores was necessary to achieve 100% agreement between the AIMS and the PDMS-2. Correlations between the PDMS-2, Locomotion and Stationary subtest and PPEDI for all age groups were in the moderate to good and good to excellent range. Non-significant correlations were found for the Reflex and Object Manipulation subtest. Conclusions: The PDMS-2 gross motor subtests were responsive to change in a heterogeneous group of children diagnosed with Pompe disease. Responsiveness concerns were identified in the Object Manipulation and Stationary subtest when the children were divided into two different functional groups. A stronger relationship was present between the PDMS-2 and AIMS age equivalent scores than the percentile scores. Motor capacity on the PDMS-2 reflects actual performance in the home and community environment as measured by the PPEDI.
Dedication

This doctoral dissertation is dedicated to my husband Troy and our children Aubrie and Georgina. Without their support and encouragement completion of this dissertation would not have been possible. I am also grateful for my mother, Myrna and for the many friends who have supported me and helped me to navigate the PhD process. Lastly I dedicate this dissertation to the children and families who live everyday with Pompe disease. They have been my best teachers.
# TABLE OF CONTENTS

Chapter I.

Introduction

<table>
<thead>
<tr>
<th>Background</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDMS-2</td>
<td>12</td>
</tr>
<tr>
<td>1. Responsiveness</td>
<td>12</td>
</tr>
<tr>
<td>2. Concurrent Validity</td>
<td>13</td>
</tr>
<tr>
<td>3. Construct Validity</td>
<td>17</td>
</tr>
</tbody>
</table>

Pompe Disease and the AGLU0 1602 and 1702 Clinical Trials | 18 |

Purpose and Implications of Research | 21 |

Operational Definitions | 24 |

References | 26 |

Chapter II.


| Introduction | 28 |
| Purpose      | 33 |
| Research Questions | 33 |
| Participants | 34 |
| Procedure    | 37 |

PDMS-2 | 38 |

Data Analysis | 39 |

Results | 40 |
Chapter III.
Concurrent Validity of the Peabody Developmental Motor Scales-Second Edition and the Pediatric Evaluation of Disability Inventory, the Pompe Pediatric Evaluation of Disability and the Alberta Infant Motor Scale

Introduction.......................................................... 56
Pompe Disease....................................................... 61
Purpose..................................................................... 63
Research Questions.................................................. 64
Participants............................................................. 65
Procedure............................................................... 66
    PDMS-2............................................................ 66
    AIMS...................................................................... 67
    PEDI and PPEDI.................................................. 68
Data Analysis.......................................................... 69

Results
    PDMS-2 and AIMS................................................ 69
    PDMS-2 and PPEDI/PEDI........................................ 70
    PPEDI and AIMS.................................................. 71
List of Tables

Chapter II.

Table 1. Gender and Age Distribution by Functional Classification... 49
Table 2. PDMS-2 Scores for Older Subjects in Functional Group ... 49
Two
Table 3. Responsiveness of PDMS-2^a Individual Subtest and ....... 50
GMC^b Percentage Scores from Baseline to Week 52; 
Functional Group One and Two Combined
Table 4. ANOVA Results for PDMS-2^a Individual Subtests and....... 50 
GMC^b Percentage Scores Factors:
Table 5. Results of Post Hoc Analysis of Locomotion Subtest by .... 51 
Time and Function^a, with Responsiveness Indices
Table 6. Results of Post Hoc Analysis of Object Manipulation ....... 51 
(OM) Subtest by Time and Function^a, with Responsiveness Indices
Table 7. Results of Post Hoc Analysis of Stationary Subtest ........ 52 
by Time and Function^a, with Responsiveness Indices
Table 8. Results of Post Hoc Analysis of Gross Motor Composite ...52 
by Time and Function^a, with Responsiveness Indices
Table 9. Example of PDMS-2^a Raw Scores at Baseline and Week...53 
52 for Two Subjects

Chapter III.

Table 1. Pearson Product Moment Correlation Coefficients (r) for.... 79 
the Age Equivalent and Percentile Scores of the 
PDMS-2 Subtests and GMC, and the AIMS
Table 2. Mean Age Equivalent and Percentile Rank Scores for....... 80 
the PDMS-2 GMC and the AIMS
Table 3. Percent Agreement in Age Equivalent Scores for the....... 80 
PDMS-2 GMC and AIMS
Table 4. Pearson Product Moment Correlation Coefficients(r)........ 81 for the PDMS-2\textsuperscript{b} Subtest Percentage Scores and the PPEDI Functional Skills Dimension Scaled Scale
Chapter I.
Introduction

Background

Documentation of eligibility for service and efficacy of intervention with standardized assessment tools is increasingly being required for service reimbursement. Eligibility for the North Carolina Infant Toddler Program currently requires children to have an established condition or developmental delay documented with scores of 2.0 standard deviations below the mean on standardized tests or 30% delay on instruments which use age equivalent test scores.\(^1\) The *Guide to Physical Therapy Practice* includes tests and measurements as an essential component in examination and evaluation, used to establish functional limitations, impairments and baseline information.\(^2\) Pediatric physical therapists are completing standardized assessments as part of their routine clinical practice with 59% reporting using a standardized measure daily or weekly.\(^3\)

In a survey of pediatric physical therapists, the Gross Motor Function Measure 88 (GMFM), the Alberta Infant Motor Scale (AIMS) and the Peabody Developmental Motor Scales-2 (PDMS-2) were the three most commonly used pediatric motor outcome measurement tools.\(^3\) The PDMS-2 was designed to assess motor development in children from birth to 72 months of age.\(^4\) It is the only discriminative outcome assessment tool of the three tools listed that
provides normative comparison data for children between 18 months and 72 months of age. The AIMS is a discriminative tool designed to assess motor development and provide normative percentile ranking from birth to 18 months of age.\textsuperscript{5} The GMFM -88 is a criterion based assessment tool that does not provide a comparison to a normative sample. The PDMS-2 measures fine motor and gross motor skills and the AIMS and the GMFM are measurements of gross motor skills only.

According to the manual, possible uses of the PDMS-2 include; “determination of motor competency relative to a normative peer sample, assessment of qualitative and quantitative capacities of individual gross motor and fine motor skills, evaluation of progress over time and determination of efficacy of interventions in research.”\textsuperscript{4} (page 8) The manual does not define appropriate use for specific diagnoses or levels of functional mobility. The PDMS-2 is widely used in a variety of clinical settings including evaluation centers for the Infant Toddler Program,\textsuperscript{1} Preschools and Schools, therapy centers and specialty clinics. It is routinely used as part of interdisciplinary clinical assessments and research clinical protocols in the Program for Neurodevelopmental Function in Rare Disease Clinic at the University of North Carolina at Chapel Hill.\textsuperscript{6} Current use of the PDMS-2 includes a very wide range of developmental disabilities and functional presentations. The available literature on PDMS-2 focuses on typically developing children, and children at risk of developmental delay. There is a paucity of evidence to support the validity of use in children with chronic disease or significant developmental disability.
The validity of a tool is not an absolute value because it varies according to the purpose of use, and the population of subjects tested. An outcome measure is valid if it measures what it is designed to measure and accurately reflects the clinical findings. Validity has also been defined as the evaluative summary of the evidence for use and the potential consequences of score interpretation. Therefore, PDMS-2 test validity should be an ongoing process in order to define appropriate use with new populations. Test validity is necessary to accurately document disease history, evaluate efficacy of the intervention, and supplement clinical decision.

Validity of an outcome measurement tool can be assessed in terms of construct, content, and concurrent validity and responsiveness to change. Responsiveness can be defined as the ability to measure clinically important change over time. The validity of a tool impacts tool responsiveness because change depends on the fit between the tool operational definition and the item you are interested in measuring. Concurrent validity is the method of evaluating criterion validity, the degree to which the instrument reflects or is related to scores obtained on a reference standard instrument. The concurrent validity between the PDMS-2 and other commonly used pediatric outcome measurement tools is inconclusive. Lack of agreement between outcome scores may change eligibility status for services between agencies and geographic regions and misinform efficacy intervention research.
The PDMS-2 is a standardized norm referenced test for children from birth to 72 months of age. Four subtests make up the PDMS-2 gross motor composite, Reflexes, Stationary, Locomotion and Object Manipulation. The Reflex subtest is indicated only for subjects less than one year of age and the Object Manipulation subtest is used only after 12 months of age. All items are scored on a three point scale, with 0 indicating that the criteria for successful performance were not met, 1 indicating the behavior is emerging, and 2 indicating successful performance of the item criteria. The PDMS-2 raw subtest scores are typically used to calculate standard percentile and age equivalent scores. The gross motor subtest standard scores are used to calculate the overall Gross Motor Quotient. PDMS-2 standard subtest scores and motor quotients are calculated using normative age data. They are beneficial for identifying risk and level of developmental delay, but are not suitable for use in determining responsiveness. Subtest raw score comparisons cannot be made because the subtests all contain a different number of test items.

**Responsiveness**

Wang et al (2006) evaluated the responsiveness to change of the PDMS 2 for children with cerebral palsy over a three-month period by use of paired t-tests, effect size, standardized response mean and the Guyatt Responsiveness Index. Subtest raw scores were used to calculate percentage scores for the subtests. Subtest percentage scores were then used to calculate gross, fine and total
motor composite percentage scores. Paired t-tests demonstrated that a statistically significant change (p < .001) was present between the gross, fine and total motor percentage scores over 3 months. The effect size for the gross motor subtest percentage score was labeled as small at 0.2 according to Cohen standards. The standardized response mean for the gross motor composite was labeled as trivial at 0.9. Information was not included on individual subtests within the fine or gross motor composites. A review of the literature did not provide any additional research on the responsiveness of the PDMS-2 gross motor subtests or composite scores.

**Concurrent Validity**

Concurrent validity has been evaluated between the PDMS-2 and other gross motor outcome measures in infants and children with typical development and risk for developmental delay. PDMS-2 test developers evaluated concurrent validity between the PDMS-2 and the Mullen Scales of Early Learning (MSEL) for typically developing children. The MSEL provides a baseline measurement of cognition and motor development and includes gross motor, visual reception, fine motor and expressive and receptive language items. The largest correlation coefficients were present between the MSEL and the PDMS-2 Gross Motor Quotient ($r = .86$) and between the MSEL and Locomotion and Object Manipulation subtests (both $r = .90$). The PDMS-2 Stationary subtest did not correlate significantly with the MSEL Gross Motor Scale.
Snyder et al (2008) examined the concurrent validity of the AIMS and the PDMS-2 gross motor subtests for infants at risk for motor delay who ranged in age from birth to eighteen months. The AIMS is an observation tool for examination of postural control in infants who range in age from birth to 18 months. The AIMS purposes are to identify infants with developmental delay, to address the rate of motor development with repeated testing and identify infants with abnormal patterns of movement. The Pearson product moment correlations varied from $r = .78$ to $.97$. The most significant correlation coefficient of $r = .97$ was found for the Locomotion subtest in infants less than nine months of age. Snyder et al (2008) used the AIMS total score and the PDMS-2 raw scores for calculation of the correlations and did not include an examination of standard scores. Eligibility for early intervention services requires providers to demonstrate developmental delay relative to a normative peer sample, using age equivalent, percentile ranking, and standard scores.

Researchers have found that the Bayley Scales of Infant Development II (BSID-II) Psychomotor Index (PDI) and the PDMS-2 had the strongest concurrent agreement for the Locomotion subtest, and that inconsistency was present in the ability to determine significant delay with standard scores. The BSID-II PDI is a discriminative motor scale that is used to identify developmental delay in children who range in age from one month to 3.5 years. Provost et al (2004) found that more than 75% of the sample classified as significantly delayed on the BSID-II (PDI $\leq 69$) did not score in the classification of “very poor” (TMQ, $=69$) on the PDMS-2. The PDMS-2 Locomotion subtest showed a high age
equivalent correlation of $r = 0.97$ with the BSID-II PDI. The Locomotion subtest and the BSID-II PDI had an age equivalent agreement within three months of 96%. The Stationary and Object Manipulation subtests and the BSID-II PDI had an age equivalent agreement within 5 months of 90-95%. With a 5-month difference in age equivalence scores obtained using different measurement tools, a child may be deemed eligible for early intervention services with one tool and ineligible for services with use of another. In a 20-month-old child, a 5-month delay in gross motor skills supports a 25% delay and eligibility for early intervention services.

Connolly et al (2006) also examined PDMS-2 concurrent validity and found low ($r = 0.30$) and non-significant correlations between the standard scores on the PDMS-2 and the BSID-II in typically developing, twelve month old infants. No correlation was found between the age equivalent scores for the Stationary and Object Manipulation subtests but a high and significant correlation was found ($r = 0.71, P < 0.05$) between the age equivalent PDMS-2 Locomotion and the BSID-II motor scale. Lack of agreement between the two tests could represent variability in service eligibility. When tests are important for clinical decision-making and service eligibility, Provost recommends a very high level of correlation ($r = 0.95$) between two tests that are used to qualify for services.

Mayrand et al (2009) examined the association between the PDMS-2 Gross Motor scale and the Pediatric Evaluation of Disability Inventory (PEDI) Functional Skills Mobility domain in children with language impairment. The PEDI is a standardized instrument for children aged 6 months to 7.5 years that uses
parent report to measure level of disability in self care, mobility and social function domains. The Functional Skills Mobility domain contains 59 items that are relevant for daily independence in mobility function such as toilet, chair, wheelchair, bed, tub, and car transfers; bed mobility; indoor and outdoor mobility and stair ascent and descent. The study found a low correlation ($r = .23$), and non-significance ($p = 0.15$) between the PEDI Mobility domain and PDMS-2 GMQ for use with children with combined physical primary language impairment and mild motor impairment. Whereas the PDMS-2 is a motor outcome measure, the PEDI focuses on the degree to which impairment impacts function in activities of daily living. Although motor capacity should reflect actual performance in ADL, Mayrand et al$^{13}$ hypothesized that the PEDI may not have been sensitive enough to detect subtle motor impairment in children with primary language impairment. The authors suggested that a stronger agreement between the PDMS-2 and the PEDI may be present if evaluated in a population of children with more significant motor impairment than the children with primary language impairment. Research supports a high intra-class correlation coefficient (.91) for concurrent validity between the PEDI and the Gross Motor Function Measure (GMFM) in children with spastic cerebral palsy.$^{14}$ The GMFM is tool designed to measure change over time in gross motor function in children with cerebral palsy.$^{14}$

Overall, when concurrent validity of the PDMS-2 is examined, literature is not available to compare the PDMS-2 to valid gross motor measurement instruments for children with a chronic disease or moderate motor impairment.$^{10}$ Wang et al found the PDMS-2 to have acceptable responsiveness to change in
the gross motor composite in children with cerebral palsy. The research did not include individual PDMS-2 subtests or investigate concurrent validity with other pediatric outcome measures. The available literature focuses on typically developing children and children at risk of developmental delay, and inconsistencies are evident with use of age equivalent scores.

**Construct Validity**

Construct validity represents the degree to which a measure reflects an operational definition. The PDMS-2 is based on the *Taxonomy of the Psychomotor Domain* by Harrow. The Locomotion subtest is designed to measure movements to change location such as rolling, crawling and walking, the Stationary subtest to measure ability to control body within center of gravity and retain equilibrium and the Object Manipulation subtest to measure coordinated motor movements such as throwing, catching and kicking a ball.

Construct validity can be reduced by item gaps or an inadequate number of items to measure the construct. The Stationary subtest includes items to measure balance in sitting and kneeling, but no items are included that capture standing balance or transitions in and out of standing. The items progress from #19, maintaining balance in kneeling for 5 seconds, to #20, standing on one foot for 3 seconds. Each item in the PDMS-2 has a normative age equivalence score recorded in months. The age equivalence score for item #19 is 13 months and for item #20 is 31-32 months. Therefore, only one item is included with an age representation between 13 and 31 months. The Locomotion subtest, however,
does contain items that measure a child’s ability to lower from standing to sitting and ability to stand independently without support. In contrast to the Stationary subtest, the AlMS provides five items that qualitatively analyze stationary standing.\(^5\)

The PDMS-2 test developers reported lower mean item discrimination coefficients for the Stationary subtest than the Locomotion subtest.\(^4\) Item discrimination, as it relates to the PDMS-2, evaluates the subtest’s ability to discriminate between different levels of function. In the Stationary subtest, the lowest discrimination coefficient \((r = .41)\) was found in the 36-47 months age group. The recommended administration Stationary subtest-starting place for the 36-47 month group is item #20, standing on one foot. A large jump in activity level is required to progress from tall kneeling to standing on one foot and item precision can be decreased. Children who do not pass the next item may have ability close to the item passed or close to the failed item, but the true level cannot be determined.\(^15\) The Stationary subtest item gaps may reduce construct validity, lead to misinterpretation of intervention efficacy with inaccurate age equivalent and standard score representation and produce low concurrent validity values when comparisons are made to other pediatric motor outcome measurement tools.

**Pompe Disease and the AGLU0 1602 and 1702 Clinical Trials**

Pompe disease is a rare lysosomal storage disease characterized by a deficiency of the enzyme acid alpha glucosidase (GAA).\(^16\) Lack of GAA causes accumulation of glycogen in cardiac, skeletal, and smooth muscle and central
nervous system tissue, leading to progressive cardiomyopathy, respiratory compromise, generalized weakness and hypotonia.\textsuperscript{16} The natural history of infantile onset Pompe disease typically leads to death by 1 year of age.\textsuperscript{17}

The Genzyme Corporation Pompe disease ERT clinical trials, AGLUO 1602 and AGLUO 1702, involved administration of ERT with recombinant human alpha glucosidase (rhGAA - Myozyme) to infants and children from birth to 42 months of age. The study’s primary objective was to evaluate the safety profile of rhGAA as determined by the proportion of patients alive and ventilator free over the course of the treatment.\textsuperscript{17} A secondary efficacy endpoint was the effect of treatment on motor development from baseline, as measured by the AIMS and or the PDMS-2.\textsuperscript{17} The Pediatric Evaluation of Disability Inventory (PEDI) and the Pompe Pediatric Disability Inventory (PPEDI) measured the change in disability index from baseline. A control group was not used. Previously, an epidemiologic study of the natural history of Pompe disease was completed to provide a historical control for the AGLUO-1602 and AGLO-1702 clinical trials.\textsuperscript{18}

Outcome measurement testing for AGLUO 1702 was completed at baseline and weeks 12, 26, 38, and 52. The AGLUO1602 cohort completed testing at baseline and weeks 12, 26, 38, 52, 64, 78, 90, and 104. Participants in the AGLUO 1702 clinical trial had the PDMS-2 completed at every assessment period. The PDMS-2 was collected for the AGLUO 1602 participants only at baseline, at the first assessment period at or after one year of age, and then at all subsequent assessments.
Numerous factors contribute to ERT response in infants and children with Pompe disease including age and severity of disease presentation, age of initiation of ERT, and GAA activity as measured by cross-reactive immunologic material (CRIM) and anti-rhGAA antibody titer levels.\textsuperscript{19} Infants who began ERT before 6 months of age, but were unable to form a native enzyme GAA, responded poorly to treatment in the trials, with no significant motor changes on the AIMS or PPEDI. Patients with a higher baseline median muscle GAA at baseline, a baseline age of less than 12 months of age, and better baseline motor scores had the best response to ERT.\textsuperscript{20}

The AGLUO1602 summary manuscript by Kishnani et al (2007) reported that 100\% of the infants survived to 18 months of age.\textsuperscript{17} Compared to the untreated historical control group, the risk of death was reduced by 99\%, and use of ventilatory assistance was reduced by 88\%.\textsuperscript{17} Thirteen of 18 children from AGLUO-1602 demonstrated motor and functional changes on the AIMS and PPEDI. At week 52, seven children demonstrated the ability to walk independently, three children could pull to stand independently and walk hand held, and three children could sit and roll independently but were not able to demonstrate weight bearing in standing.\textsuperscript{17} The remaining five subjects did not demonstrate clinically meaningful change on the AIMS or PPEDI.\textsuperscript{17} The children with the largest change scores on the AIMS also demonstrated the most substantial functional gains on all three PPEDI domains, self-care, mobility and socialization.\textsuperscript{17}
Sixteen of 21 children were alive at the end of 52 weeks in the AGLUO-1702 clinical trial. \textsuperscript{21} Five patients died before week 28 due to cardiac or respiratory failure. \textsuperscript{21} At baseline, five children were invasively ventilated and two were noninvasively ventilated (via mask). Of the five children who were invasively ventilated, three remained ventilated 24 hours a day, one reduced ventilation to 12 hours a day and one child died. Nicolino et al (2009) documented an age equivalent motor change score in 13 of 21 children on the AIMS or PDMS-2, but no scores or data analyses were included in the manuscript for interpretation of change scores. \textsuperscript{21} The authors reported that five children were able to walk independently and eight patients could sit independently. \textsuperscript{21} The remaining eight patients made no significant motor development gains from baseline.\textsuperscript{21}

**Purpose and Implications of this Research**

The first purpose of this dissertation was to examine the responsiveness of the gross motor subtests of the PDMS-2, in children diagnosed with Pompe disease who participated in Enzyme Replacement Therapy (ERT). Responsiveness was examined by level of function in two groups of children diagnosed with Pompe disease; children who ambulate independently and children who require use of assistive devices to ambulate or are unable to ambulate. The second purpose was to examine the concurrent validity between the PDMS-2 and the AIMS, the PEDI, and the PPEDI in children diagnosed with Pompe disease.
The dissertation involved secondary data analyses of children diagnosed with Pompe disease, from the Genzyme Corporation enzyme replacement therapy (ERT) clinical trials. The PDMS-2 was concurrently administered with the AIMS, PEDI and PPEDI to 18 subjects under the age of 7.2 months (AGLUO - 1602) and 21 children between 6 months and 42 months (AGLUO 1702). The PDMS II was collected but never analyzed in clinical trial manuscripts.

Motor gains in response to ERT have been documented in the literature without a defined level of meaningful change response. The ALGUO 1602 and 1702 data sets provided an opportunity to measure PDMS-2 gross motor responsiveness in children diagnosed with Pompe disease who were diverse in age and had a heterogeneous level of functional mobility. Only children who were documented as “motor responders” in the literature were included in the analyses.

The author hypothesized that the Stationary subtest would be less responsive than the Locomotion or Object Manipulation subtest and that responsiveness would be dependent on level of functional mobility. Lack of responsiveness was hypothesized in the group of children who have difficulty maintaining balance in standing, have delayed acquisition of independent walking and may require use of adaptive equipment or caregiver assistance to maintain standing or ambulate.

The AIMS and the PDMS-2 both provide normative scores and are commonly used in early intervention to determine eligibility for services.
Variability in normative scores creates challenges in test interpretation for eligibility determination and interagency collaboration. Examination of concurrent validity in a sample of children diagnosed with Pompe disease may aid therapists in test interpretation and identification of tool strengths and limitations that are applicable to a larger range of developmental disabilities. The author hypothesized that the strongest PDMS-2 subtest relationship would be present between the Locomotion subtest and the AIMS.

A comprehensive evaluation for children diagnosed with Pompe disease should not be limited to assessments that evaluate capacity to perform motor skills in the clinic environment. Function should also be considered within the context of activity limitations and participation restrictions within the community, home, and school environment. This research provided an opportunity to determine whether motor capacity on the PDMS-2 and the AIMS reflects actual performance in activities of daily living on the PEDI and the PPEDI. The author hypothesized that the PPEDI would have a stronger relationship with the PDMS-2 than the PEDI and that the strongest PDMS-2 subtest relationship would be between the Locomotion subtest and the PEDI and the PPEDI Functional Skills Mobility dimension.
Operational Definitions:

1. **Validity** will be defined as the evaluative summary of the evidence for use and the potential consequences of score interpretation. Validity of a tool is not an absolute value because it varies according to the purpose of use, and the population of subjects tested.

2. **Construct validity** indicates the degree to which a measurement reflects an operational definition. The construct validity of an instrument may evolve over time with changes in the understanding of the construct of interest.

3. **Content validity** defines the extent to which the items reflect a content domain and it is clear that no items are missing.

4. **Concurrent validity** is the method of evaluating criterion validity, the degree the instrument reflects or is related to scores obtained on a reference standard instrument. Concurrent validity also evaluates the level of agreement that is present between two outcome measurements.

5. **Responsiveness** can be defined as the ability to measure clinically important change over time.

6. **Minimal detectable change** has been described as the amount of change that exceeds the standard error of the instrument.

7. **AGLUO 1702**: Genzyme Corporation Clinical Trial for Myozyme-Alglucosidase alfa: recombinant human alpha glucosidase (rhGAA). Clinical trial for children from 6 to 36 months of age at baseline. Two children were included in the trial with baseline ages of > 36 months. AGLUO 1702 completed enrollment first and then AGLUO 1602 was initiated.

8. **AGLUO 1602**: Genzyme Corporation Clinical Trial for Myozyme-Alglucosidase alfa: recombinant human alpha glucosidase (rhGAA). Clinical trial designed for children from birth to 6 months of age at baseline. One child was included in the trial with a baseline age of 7.2 months.

9. **ERT**: enzyme replacement therapy

10. **GMQ**: The Gross Motor Quotient is derived from the standard scores of three gross motor subtests. A total gross motor function standard score is determined from norm-referenced data. It measures a child’s total gross motor development. A quotient of 80-89 is considered “below average” and a 70-79 score “poor”.


11. **GMC**: Gross Motor Composite. Each gross motor subtest raw score is transformed into a percentage score. The subtest percentage scores are summed and divided by three to create the GMC.

The data set will be described in two manuscripts.


**Aims:**

a) To investigate the responsiveness to change of the PDMS-2 in children with Pompe disease who participated in ERT and demonstrated clinically meaningful change on the PPEDI and AIMS.

b) To investigate the responsiveness to change of the PDMS-2 in children with Pompe disease who participated in ERT when subjects are classified according to level of functional mobility.


**Aim:**

To evaluate the concurrent validity of the PDMS-2 and the Pediatric Evaluation of Disability Inventory, the Pompe Pediatric Evaluation of Disability and the Alberta Infant Motor Scale in a sample of infants and children with Pompe disease
References

1. Services NCEI. New Eligibility Definition For the NC Infant Toddler Program http://www.ncei.org/ei/.


Chapter II.

Introduction

As medical advances continue to increase life expectancy in rare diseases such as Pompe disease, valid assessment of developmental motor outcomes is necessary to accurately document disease history, evaluate efficacy of the intervention, and supplement clinical decision-making. The validity of a tool is not an absolute value because it varies according to the purpose of use, and the population of subjects tested.¹ Test validity should be an ongoing process in order to define appropriate use with new populations. When investigators are faced with choosing outcome measures, in the absence of a gold standard, test selection is limited to the best available measure. They look to tools with good psychometric properties for the defined research age group that have been used in other diagnoses with similar neuromuscular presentations. Researchers must then continue to evaluate new evidence for application to their clinical population.¹

A motor outcome measure with adequate responsiveness to change has not been established for use in infantile Pompe disease. Pompe disease is a rare lysosomal storage disease characterized by a deficiency of the enzyme acid alpha glucosidase (GAA).² Lack of GAA causes accumulation of glycogen in
cardiac, skeletal, and smooth muscle and central nervous system tissue, leading to progressive cardiomyopathy, respiratory compromise, generalized weakness and hypotonia.\(^2\) The natural history of infantile onset Pompe disease typically leads to death by one year of age.\(^3\) Infants treated with recombinant human acid \(\alpha\) glucosidase (Myozyme) have been shown to have decreased ventilator use and decreased risk of death. Developmental motor skill acquisition in response to Myozyme has been variable.

Motor efficacy was measured in the pivotal Genzyme Corporation, Myozyme clinical trials (AGLUO 1602 and AGLUO 1702) with administration of the Peabody Developmental Motor Scales- second edition (PDMS-2), the Alberta Infant Motor Scale (AIMS), the Pediatric Evaluation of Disability Inventory (PEDI) and the Pompe Pediatric Evaluation of Disability Inventory (PPEDI).\(^3\) Kishnani et al (2007) reported that 13 of 18 children in AGLUO-1602 demonstrated motor and functional changes on the AIMS and PPEDI.\(^3\) Nicolino et al (2009) documented an “age equivalent motor change score” in AGLUO-1702, but no scores or data analyses were included in the manuscript.\(^4\) Functional skill acquisition varied from the ability to roll and sit independently to walking independently.\(^3\) No functional outcome data have been published on the PDMS-2 motor outcomes.

According to the manual, possible uses of the PDMS-2 include; “determination of motor competency relative to a normative peer sample, assessment of qualitative and quantitative capacities of individual gross motor and fine motor skills, evaluation of progress over time and determination of
The manual does not indicate that the PDMS-2 is most appropriate for specific diagnoses or defined levels of functional mobility. The PDMS-2 is widely used in a variety of clinical settings including evaluation centers for the Infant Toddler Program, preschools and schools, therapy centers and specialty clinics. Therefore, current use includes children with a very wide range of developmental disabilities and functional presentations. There is a paucity of evidence to support the validity of use in children with chronic disease or significant developmental disability. A review of the literature identified only one article on the responsiveness of the PDMS-2 gross motor subtests or composite scores.

Responsiveness can be defined as the ability to measure clinically important change over time. Wang et al (2006) evaluated the responsiveness to change of the PDMS-2 in children with cerebral palsy over a three-month period using paired t-tests, effect size, standardized response mean and the Guyatt Responsiveness Index. Subtest raw scores were used to calculate percentage scores for the subtests. Subtest percentage scores were then used to calculate gross, fine and total motor composite percentage scores. Paired t-tests demonstrated that a statistically significant change (p<. 001) was present from baseline to end of study in the gross, fine and total motor percentage scores. The effect size for the gross motor subtest percentage score was labeled as small at 0.2 according to Cohen standards. The standardized response mean for the gross motor composite was labeled as trivial at 0.9. Information was not included on individual subtests within gross motor composites.
Responsiveness depends on the number of items on the instrument and the standard error of measurement of the instrument.\textsuperscript{10} The more items on the instrument, and the smaller the standard error of measurement, the greater the opportunity to detect change.\textsuperscript{10} Content validity defines the extent to which the items reflect a content domain and it is clear that no items are missing. According to the PDMS-2 manual the Stationary content measures “the child’s ability to sustain control of his or her body within its center of gravity and retain equilibrium.”\textsuperscript{4} (page 34) The Stationary subtest may have reduced content validity and reduced responsiveness due to a lack of inclusion of items to measure standing balance. The Stationary items progress from maintaining balance in kneeling for five seconds to standing on one foot independently with no items available to qualitatively analyze stationary standing or transitions in and out of standing. Only one item is included with an age representation between 13 and 31 months. Instrument precision can be decreased if large jumps in activity level are required to gain item credit. Children who do not pass the next item may have ability close to the item passed or close to the failed item but the true level cannot be determined.\textsuperscript{11} The Stationary subtest item gaps may lead to decreased test responsiveness in children who have difficulty with standing balance, have delayed acquisition of independent walking, or require use of adaptive equipment or caregiver assistance to maintain standing or to ambulate.

The PDMS-2 test developers report content validity with mean item discrimination coefficients in each subscale.\textsuperscript{5} In the Stationary subscale the
lowest discrimination coefficient, $r = .41$ was found with children in the normative sample in the 36-47 month age group. Item #20, standing on one foot for 3 seconds is the recommended administration starting place for children 27-38 months of age.

Stationary subtest item gaps may lead to misinterpretation of intervention efficacy and inaccurate age equivalent and standard score representation. Gross motor subtest raw scores are converted to subtest standard scores and combined to form the Gross Motor Quotient (GMQ). Test developers recommend the most credence be given to the motor quotient scores when important decisions are to be made about diagnosis or placement. All standard scores are calculated based on the number of credited items in the normative sample. A Stationary subtest raw score that has reached a plateau could present as a lower standard score at subsequent testing, and result in a lower and possibly inaccurate GMQ.

Research supports that the psychometric properties on the PDMS-2 are poorer for the Stationary than the Locomotion subtest. In examining the concurrent validity of the PDMS-2 with the AIMS and the Bayley Scales of Infant Development-II (BSID-II), the largest subtest coefficient correlations were found with the Locomotion subtest. Snyder et al (2008) examined the concurrent validity of the AIMS and the PDMS 2 gross motor subtests in infants at risk for motor delay from birth to eighteen months of age. The Pearson product moment correlations varied from $r = .78$ to .97. The largest correlation coefficient of $r = .97$ was found for the Locomotion subtest in infants less than nine months of
age. Concurrent validity has also been evaluated between the PDMS-2 and the BSID-II in typically developing, twelve month old infants.\textsuperscript{14} The standard scores on the PDMS-2 GMQ and the BSID-II Motor Scale showed a low and non-significant correlation value of $r = 0.30$. No correlation was found between the age equivalent scores for the Stationary and Object Manipulation subtests and the BSID-II, but a high and significant correlation ($r = 0.71\ p < 0.05$) was found between the age equivalent PDMS-2 Locomotion subtest and the BSID-II.

**Purpose**

The purpose of this study was to examine the responsiveness of the PDMS-2 in an investigation of intervention efficacy for children diagnosed with Pompe disease. Secondary analyses of the PDMS-2 data, from the Genzyme Corporation, Myozyme, enzyme replacement therapy (ERT) clinical trials were completed. The responsiveness to change on the PDMS-2 individual gross motor subtests was examined in children who had been diagnosed with Pompe disease and who demonstrated clinically meaningful change on the AIMS and PPEDI. A second purpose was to compare responsiveness to change in two groups of children diagnosed with Pompe disease who had different levels of functional mobility.

**Research Questions**

1. Are the PDMS-2 gross motor subtests and gross motor composite score responsive to gross motor change over a 52-week period in children
diagnosed with Pompe disease who demonstrated clinically meaningful change on the AIMS and PPEDI?

2. Are the PDMS-2 gross motor subtests and gross motor composite score more responsive to change in motor function in children diagnosed with Pompe disease who ambulate without assistive devices, as compared to children who ambulate with an assistive device or who are unable to ambulate?

It was hypothesized that the Stationary subtest would be less responsive than the Locomotion or Object Manipulation subtest. Lack of responsiveness was expected to be most problematic for children with decreased dynamic standing balance and delayed acquisition of independent walking. These data analyses may help to inform clinical decision making for children with Pompe disease and design of future clinical trials. The children in these data sets are diverse in age and functional presentation. The functional presentations are similar to many children who are typically followed in pediatric clinics. Identification of PDMS-2 strengths and limitations for use may be clinically applicable for a larger range of developmental disabilities.

Participants

Fourteen of the 34 children from AGLUO-1602 and 1702 data were included in the analyses. Twenty-three of the 34 children had clinically significant change on the PPEDI Functional Skills Mobility dimension and improvements in the total AIMS score at the one-year follow-up. Nine of the 23 were excluded for
the following reasons; six subjects did not have complete PDMS-2 data for the testing time frames, two children had a cross reactive immunologic negative status (CRIM-) and one patient showed improvement initially and then showed decline at the final assessment. Total absence of GAA is described as CRIM – and Myozyme ERT has been shown to be less effective in CRIM- patients.³

A level of clinically meaningful change for the AIMS and the PPEDI has not been established in children diagnosed with Pompe disease. Minimal detectable change has been described as the amount of change that exceeds the standard error of the instrument.¹⁰ The test developers for the PEDI and PPEDI recommend use of the standard error in interpretation, and state that change that exceeds 2x the standard error represents improved functional performance.¹⁵ Clinically meaningful change for this study was defined as change scores that exceeded 2x the standard error.

Functional Classification

The FDA submission for Myozyme market approval and manuscripts summarizing the clinical trials have classified responders as independent walkers, standers/walkers with assistance and independent sitters who were unable to sustain weight bearing in standing.⁴,¹⁶ Similarly, the children in this analysis were divided into three groups according to level of functional mobility. The AIMS scores at one year follow-up determined functional group placement.

**Group one** included children who could transition in and out of standing without adult assistance and walk without an assistive device if they were older than two years of age. If they were 18
months to two years of age they were able to cruise independently at furniture and stand momentarily without assistance. Inclusion was determined by an AIMS standing score of 13 or more for children at least two years of age and an AIMS standing score of 10 or more for children younger than two years of age. Six children were classified as group one.

**Group two** included children who sat independently and were able to quadruped crawl. They were able to take weight into their legs and stand or walk with use of assistive devices or caregiver. An AIMS standing score of 3-9, AIMS sitting score of 10 or greater and a prone score of 17 or greater determined inclusion. Only one child was classified as group two.

**Group three** included children who were able to sit independently. If mobility was present, it was through belly crawling or wheelchair propulsion. Inclusion was determined by AIMS sitting score of less than 10. Seven children were classified as group three.

Group two and three were combined to form one functional group for data analyses because only one subject was classified as group two. The combined group was labeled as group two and included all children who did not demonstrate the ability to walk or cruise independently.

Table 1 includes a gender and age distribution by functional classification. Group two had an older group mean age (21.5 months) than group one (15
months). Group two included a 37- and a 43-month-old child who fell outside of the original clinical trial protocol inclusion criteria of 36 months. Functional presentation in Pompe disease is not assumed to improve linearly by age. The older subjects in group two did not have the lowest or highest function in the group when subjects were ranked from lowest to highest function with Gross Motor Composite (GMC) percentage scores. Subject A ranked 3/8 and subject B ranked 5/8. The ranking indicates that although they were the oldest they did not have the lowest or highest functional skills in group two. Table 2 displays the older children’s individual subtest percentage scores and the mean group two subtest scores.

**Procedure**

The AGLUO-1602 and 1702 data were collected by physical therapists in the USA, Taiwan, Europe and Israel. The primary author of this manuscript collected USA clinical data and assisted with international reliability training. Reliability training included a review of the theoretical basis for each assessment, instruction on administration guidelines and data reporting and establishing inter-rater reliability of 90% agreement for all of the tests. To minimize investigator bias, the assessors did not review previous scores prior to administration of a repeat assessment. The same therapist at each site administered all repeat assessments. Following administration of the assessment score sheets were forwarded to central scoring and entered into a database.
The secondary analyses first included identification of children who demonstrated clinically meaningful change scores. Then the identified children were classified into functional group one and two for statistical analyses.

PDMS-2

The PDMS-2 is a standardized norm referenced test for children from birth to 72 months of age. Four subtests make up the gross motor composite, Reflexes, Stationary, Locomotion and Object Manipulation. All items are scored on a three point scale with 0 indicating that the criteria for successful performance were not met, 1 indicating the behavior is emerging and a 2 indicating successful performance of the item criteria. The PDMS-2 raw subtest scores typically are used to calculate standard percentile and age equivalent scores. The gross motor subtest standard scores are used to calculate the overall Gross Motor Quotient. The Reflex subtest is indicated only for subjects less than one year of age, and Object Manipulation subtest is used only after 12 months of age. PDMS-2 standard subtest scores and motor quotients are calculated using normative age data. They are beneficial for identifying risk and level of developmental delay but are not suitable for use in determining responsiveness. Subtest raw score comparisons cannot be made because the subtests all contain a different number of test items. Similar to PDMS-2 responsiveness work completed by Wang et al, (2006), each subtest raw score was transformed into a percentage score. The percentage score was calculated by dividing the obtained raw score by the maximum raw score and multiplying by 100.
The PDMS-2 was administered according to the guidelines in the manual. The Stationary, Locomotion and Object Manipulation subtest percentage scores were used to calculate the Gross Motor Composite (GMC) score because all participants were greater than one year of age at the established baseline entry period. The original protocol specified that the PDMS-2 would be completed at study initiation, the first assessment after one year of age, and then every 3 months until study completion. The first assessment period after one year of age was defined as baseline for this analysis, and the second assessment was 52 weeks after the established baseline. The protocol allowed a +/- 7day window for assessment administration. A baseline entry of at least 12 months of age allowed subjects to be analyzed on the same three subtests, Stationary, Locomotion and Object Manipulation. It also allowed the same three subtests to be used to calculate the GMC percentage scores. An entry assessment age of at least 12 months also captures the developmental time frame in which standing stationary and transition skills are expected to begin emerging. Use of this time frame was essential to evaluate the responsiveness concerns in the Stationary subtest.

**Data Analysis**

Descriptive statistics were calculated for the group as a whole and for children in each of the functional groups (Table 1). Responsiveness to change in the individual subtests and the GMC in the whole group was calculated with four one way repeated measure ANOVAs. Subtest and GMC responsiveness by function was calculated with four two way repeated measure ANOVAs. Each ANOVA had two factors, time (within group at baseline and week 52) and
functional classification (between group one and two). A SAS general linear model framework was used for all ANOVAs. A Tukey–Kramer correction factor was used to adjust for multiple comparisons at an a priori significance level of 0.05.

Two response indices were calculated, Effect Size (ES) and Standardized Response Mean (SRM). ES is the measure of change obtained by dividing the mean change from baseline to week 52 by the pooled standard deviation of baseline and week 52. Absolute standards for interpretation of ES are not available. The values of the ES were interpreted according to Cohen recommendations as trivial (ES of <0.2), small (ES of >=0.2 <0.5), moderate (ES of >=0.5<0.8), or large (ES of >=0.8) A small effect size is 20% of one standard deviation.\textsuperscript{9} SRM provides an estimate of change in the measure, standardized relative to between-subject variability in change scores.\textsuperscript{9} SRM is ES adjusted for the value of the correlation coefficient between the baseline and week 52 mean values.\textsuperscript{17} Cohen standards were used for interpretation.\textsuperscript{9}

Results

Summary results are presented in Table 3 for the whole group with functional groups one and two combined. Repeated measure ANOVAs documented that a significant difference was present between the mean subtest percentage values at baseline and at week 52. The ES and SRM were moderate for the Locomotion subtest and large for the Object Manipulation and Stationary subtests and the GMC.
Table 4 outlines individual subtest and GMC responsiveness by the factors of Time and Function. Although an interaction effect was found in the Object Manipulation subtest, it was not relevant to the research questions in this manuscript. The test for simple main effects for Time and Function found a significant difference in the mean percentage scores for all three subtests and for the GMC.

**Locomotion**

Table 5 outlines the post hoc analysis and the response indices for the Locomotion subtest. A significant difference was present in the mean percentage scores from baseline to week 52 for both functional groups. A significant difference between the two functional groups was not present at baseline, but was present at week 52. The ES and SRM values were large for functional group one and moderate for functional group two.

**Object Manipulation**

Table 6 outlines the post hoc analysis and the response indices for the Object Manipulation subtest. A significant difference was present in the mean percentage scores from baseline to week 52 for functional group one only. A significant difference between functional group one and two was not present at baseline, but was present at week 52. The ES and SRM values were large for functional group one and small for functional group two.
Stationary

Table 7 outlines the post hoc analysis and the response indices for the Stationary subtest. No significant difference was present between baseline and week 52 mean percentage scores for either functional group. In addition, no significant differences were present between groups at either time point. The ES and SRM values were large for both functional groups.

GMC

Table 8 outlines the post hoc analysis and the response indices for the GMC. A significant difference was present between baseline and week 52 mean percentage scores for both functional groups. No significant differences between groups were present at either time point. The ES and SRM values were large for functional group one and moderate for functional group two.

Discussion

The results of this study support adequate responsiveness of the PDMS-2 percentage GMC scores in a heterogeneous group of children with Pompe disease. Wang et al (2006) also documented acceptable responsiveness of the composite scores of the PDMS-2 in children with cerebral palsy with a wide range of functional abilities. However, Wang did not measure responsiveness separately for varied functional presentations and for each separate subtest. The mean GMC percentage scores were significantly different from baseline to week 52 in both functional groups. The large Locomotion subtest mean percentage change score in group two could have influenced the overall GMC
percentage change score because a significant difference was not found in the other two subtests. The PDMS-2 manual recommends the GMC as the best estimator of gross motor competency, and the composite score is derived from all three subtests.  

Although the testing time frame for this research was 52 weeks, clinical reevaluations for reauthorization of therapy services often need to be completed every three to six months. Over this shorter time frame, a smaller change in Locomotion subtest score may be present. A smaller Locomotion subtest change score, paired with the same lack of change on the other two subtests, may not produce a significantly different mean GMC change score.

“Responsive measures discriminate between trivial and substantial change within groups and differentiate between those groups.”  

The Locomotion subtest was able to discriminate change within both groups from baseline to week 52 and between groups at both time periods. The Stationary subtest was unable to discriminate within group or between group differences. The Object Manipulation subtest discriminated change within group one from baseline to week 52 and between the functional presentations of group one and group two at week 52. The Locomotion subtest measures the child’s ability to move his or her body from one base of support to another and includes not only ambulation, but rolling, pivot in sitting, and crawling. Although functional group one and two had different levels of functional locomotion, the Locomotion subtest was responsive to change in both groups. An essential component of Locomotion is balance. The Stationary subtest measures balance and is defined in the
manual as “the ability to sustain body control within the center of gravity and retain equilibrium.” 4 (page 34) Gross motor skill acquisition should be evident not only in the Locomotion subtest but paired to some extent with change in the Stationary and Object Manipulation subtests. Items are present on the Stationary subtest to capture balance in sitting, but no items are present to quantify quality of bipedal standing balance or transitions in and out of standing. Table 9 shows an example of PDMS-2 scores for two subjects, one from each functional group. The children had stationary subtest raw scores that were very similar (38, 36), but Locomotion subtest raw scores that were very different (113, 44). The Locomotion raw subtest scores equate to very different functional presentations. Subject A was able to ambulate independently and walk up stairs, and had emerging jumping skills. Subject B could sit independently, pivot in sitting, and creep on hands and knees. The Stationary subtest scores do not differentiate between the children’s different levels of balance.

Responsiveness and content validity are dependent on an adequate number of test items to capture the content domain and measure meaningful change. Thirteen out of fourteen children had a Stationary subtest week 52 raw score between 36 and 38. This represents a ceiling effect prior to either item #19, maintaining balance in tall kneeling or item #20 standing on one foot for 3 seconds.

The Object Manipulation subtest was responsive to change within group one for children who have the ability to stand independently and throw or kick a ball. Children with a greater degree of developmental disability with an inability to
maintain independent standing could not progress on the Object Manipulation items (functional group two). Item #1 and #2 require only independent sitting balance to catch and roll a ball. But item #3 requires independent standing to fling a ball. The children in functional group two were limited in their ability to successfully complete items beyond #2. Five of the eight children in functional group two had a final score on the Object Manipulation subtest of 4 or less. Poor responsiveness to change in functional group two was supported with a non-significant ANOVA value and a small SRM of .47. The Object Manipulation subtest for functional group one had a large effect with a SRM of 3.8.

Response indices values indicate that the Locomotion and the GMC were also more responsive to change in functional group one than group two. Values were in the large range for functional group one and in the moderate range for functional group two. The ES is calculated by dividing the mean change score by the pooled standard deviation of the baseline and week 52 scores. Functional group one had a larger mean change score and less variance in scores than functional group two, thus creating larger response indices. The Stationary subtest ES and SRM values were large in both functional groups and did not support the lack of responsiveness to change found in the post hoc p-values. Effect size is affected by group variance. The Stationary subtest had a very small score variance, thus creating a large effect size. Therefore the ES and SRM should be interpreted with caution because the lack of items to capture standing function actually created a large ES.
Application to Clinical Practice and use in Research

Although the results of this research are directly applicable only to children with a diagnosis of Pompe disease, the identified limitations of use may be appropriate to consider with other diagnoses. The functional distribution of the sample was not dissimilar to what might be seen in typical pediatric clinic environments. Children with other diagnoses can present similar to functional group one with generalized weakness, delayed acquisition of independent ambulation and challenges with high balance and coordination skills or similar to functional group two with the ability to sit independently, required use of adaptive equipment to ambulate or stand, and primary community mobility with a wheelchair. Therapists and physicians should exercise caution in interpretation of the Stationary subtest as a valid evaluative measure of change. Similarly, the Object Manipulation subtest may not measure change in function for children with limited ability to maintain independent standing. Inclusion of subtest scores on reports should be paired with a detailed qualitative description of functional presentation and areas of noted decline or improvement. The Locomotion subtest may be the most useful test for evaluative measurement in clinical and research protocols.

Conversion of subtest raw scores into percentage scores may be beneficial for interpretation as an evaluative measure. Norm referenced standard scores are appropriate for use as a discriminative tool but very challenging to interpret as an evaluative measure of change. Accurate standard score interpretation by health care and school professionals, insurance agencies and
parents is essential to guide decision making and plan of care development. Professionals and parents may be unaware that the same standard score on two consecutive testing periods actually represents motor skill acquisition. The raw scores in this research were easily converted into percentage scores. PDMS-2 percentage score use as a measurement of change in performance would be very similar to the percentage scores on the GMFM 88 or the scaled scores on the PEDI.\textsuperscript{15,18}

This research highlighted the importance of functional classification of subjects in Pompe disease efficacy research. Ongoing development of a more detailed functional classification system for use in Pompe disease is recommended. The PDMS-2 may be most appropriately used in children with mild developmental delay, who have the capacity or potential to; balance in standing to complete ball skills and acquire higher standing skills like standing on one foot. Use of the Locomotion subtest as a discriminative measure to define function relative to a normative sample may also be appropriate. A wide heterogeneity is present in the functional presentation of children with Pompe disease, and functional course is not always predictable. For a study sample that includes a majority of children with significant motor impairment, the Stationary and Object Manipulation subtests may not have adequate responsiveness to change.
Limitations

Despite the sample size, this study did serve to identify some potential areas of concern with the use of the PDMS-2 gross motor scales in children with Pompe disease. Pompe disease is an extremely rare condition, and recruitment was limited to a convenience sample. Sample sizes for the secondary analyses were also limited due to incomplete data at some assessment periods.

Conclusion

The PDMS-2 manual provides no guidelines for interpretation of standard scores as evaluative measures. Conversion of PDMS-2 subtest raw scores into percentage scores supported increased utility as an evaluative measure.

The PDMS-2 gross motor subtests and GMC percentage scores were responsive to change in a heterogeneous group of children with Pompe disease. Responsiveness concerns were identified in the Object Manipulation and Stationary subtest when the children were divided into two different functional groups. Level of functional mobility should be taken into consideration when evaluating PDMS-2 use for future Pompe disease research.

The Locomotion subtest may be the most useful PDMS-2 gross motor subtest for evaluative measurement of clinical change in Pompe disease. The Locomotion subtest had adequate within-group responsiveness over time and was able to discriminate between the two groups at week 52. The Responsiveness to change results on the Stationary subtest were inconsistent between the post hoc analyses and the response indices. Additional study is
justified to explore the clinical and research utility of the PDMS-2 in this population.

Table 1.

Gender and Age Distribution by Functional Classification

<table>
<thead>
<tr>
<th>Function</th>
<th>n</th>
<th>Gender</th>
<th>Mean Age at Baseline in months</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group One</td>
<td>6</td>
<td>4 M, 2 F</td>
<td>15.0</td>
<td>2.5</td>
<td>12</td>
<td>19</td>
</tr>
<tr>
<td>Group Two</td>
<td>8</td>
<td>5 M, 3 F</td>
<td>21.5</td>
<td>11.1</td>
<td>12</td>
<td>43</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>9 M, 5 F</td>
<td>18.9</td>
<td>9.0</td>
<td>12</td>
<td>43</td>
</tr>
</tbody>
</table>

Table 2.

PDMS-2a Percentage Scores for Older Subjects in Functional Group Two

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Stationary</th>
<th>Locomotion</th>
<th>OMb</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. 37 months of age</td>
<td>53</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>B. 43 months of age</td>
<td>60</td>
<td>25</td>
<td>8</td>
</tr>
<tr>
<td>Functional Group Two</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Percentage Scores</td>
<td>50.2(12.4)</td>
<td>20.8(18.1)</td>
<td>9.6(12.3)</td>
</tr>
</tbody>
</table>

a PDMS-2: Peabody Developmental Motor Scales- 2nd edition

b OM - Object Manipulation
Table 3.
Responsiveness of PDMS-2\textsuperscript{a} Individual Subtest and GMC\textsuperscript{b} Percentage Scores from Baseline to Week 52; Functional Group One and Two Combined

<table>
<thead>
<tr>
<th></th>
<th>Baseline Mean (SD)</th>
<th>Week 52 Mean (SD)</th>
<th>F value (p)</th>
<th>ES\textsuperscript{c}</th>
<th>SRM\textsuperscript{c}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locomotion</td>
<td>30.2(17.4)</td>
<td>42.6(19.1)</td>
<td>26.71(&lt;.0001)*</td>
<td>0.68</td>
<td>0.58</td>
</tr>
<tr>
<td>OM\textsuperscript{d}</td>
<td>13.0(11.2)</td>
<td>31.3(21.5)</td>
<td>22.86(.0004)*</td>
<td>1.07</td>
<td>1.05</td>
</tr>
<tr>
<td>Stationary</td>
<td>56.2(8.3)</td>
<td>62.3(4.2)</td>
<td>8.71(.011)*</td>
<td>0.92</td>
<td>0.87</td>
</tr>
<tr>
<td>GMC\textsuperscript{b}</td>
<td>34.2(11.3)</td>
<td>46.7(13.9)</td>
<td>38.02(&lt;.0001)*</td>
<td>.98</td>
<td>0.94</td>
</tr>
</tbody>
</table>

\textsuperscript{a}PDMS-2: Peabody Developmental Motor Scales- 2\textsuperscript{nd} edition

\textsuperscript{b}GMC-Gross Motor Composite

\textsuperscript{c}Cohen’s d used for Effect Size (ES) and Standardized Response Mean (SRM); trivial <0.2, small =0.2 <0.5, moderate >=0.5<0.8, large >=0.8

\textsuperscript{d}OM - Object Manipulation

Table 4.
ANOVA Results for PDMS-2\textsuperscript{a} Individual Subtests and GMC\textsuperscript{b} Percentage Scores Factors:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Locomotion F value (p)</th>
<th>OM\textsuperscript{c} F value (p)</th>
<th>Stationary F value (p)</th>
<th>GMC\textsuperscript{b} F value (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>31.97(.0001)*</td>
<td>72.04(&lt;.0001)*</td>
<td>10.13(.008)*</td>
<td>4.75(&lt;.001)*</td>
</tr>
<tr>
<td>Function</td>
<td>10.74(.0066)*</td>
<td>7.06(.021)*</td>
<td>6.34(.027)*</td>
<td>4.33(.059)</td>
</tr>
<tr>
<td>Time*Function</td>
<td>2.61(.1320)</td>
<td>23.30(.0004)*</td>
<td>.31(.590)</td>
<td>2.85(.117)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}PDMS-2: Peabody Developmental Motor Scales- 2\textsuperscript{nd} edition

\textsuperscript{b}GMC-Gross Motor Composite

\textsuperscript{c}OM-Object Manipulation
### Table 5.

**Results of Post Hoc Analysis of Locomotion Subtest by Time and Function\(^a\), with Responsiveness Indices**

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Function</th>
<th>Baseline Mean Percentage (SD)</th>
<th>Week 52 Mean Percentage (SD)</th>
<th>t value (adj. p)</th>
<th>ES(^b)</th>
<th>SRM(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locomotion</td>
<td>1</td>
<td>41.7(9.6)</td>
<td>58.2(6.5)</td>
<td>5.36(.0009)*</td>
<td>2.02</td>
<td>2.66</td>
</tr>
<tr>
<td>Locomotion</td>
<td>2</td>
<td>20.8(18.1)</td>
<td>32.9(14.0)</td>
<td>4.51(.003)*</td>
<td>0.75</td>
<td>0.66</td>
</tr>
<tr>
<td>t value (adj. p)</td>
<td></td>
<td>2.88(.059)</td>
<td>3.49(.02)*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Post Hoc analyses with Tukey-Kramer test for multiple comparisons

\(^b\)Cohen’s d used for Effect Size(ES) and Standardized Response Mean(SRM); trivial <0.2, small =0.2 <0.5, moderate >=0.5<0.8, large >=0.8

### Table 6.

**Results of Post Hoc Analysis of Object Manipulation (OM) Subtest by Time and Function\(^a\), with Responsiveness Indices**

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Function</th>
<th>Baseline Mean Percentage (SD)</th>
<th>Week 52 Mean Percentage (SD)</th>
<th>t value (adj. p)</th>
<th>ES(^b)</th>
<th>SRM(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OM</td>
<td>1</td>
<td>17.0(9.2)</td>
<td>48.3(14.9)</td>
<td>8.81(&lt;.0001)*</td>
<td>2.53</td>
<td>3.80</td>
</tr>
<tr>
<td>OM</td>
<td>2</td>
<td>9.6(12.3)</td>
<td>18.0(16.8)</td>
<td>2.80(.067)</td>
<td>0.57</td>
<td>0.47</td>
</tr>
<tr>
<td>t value (adj p)</td>
<td></td>
<td>97(.768)</td>
<td>4.10(.007)*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Post Hoc analyses with Tukey-Kramer test for multiple comparisons

\(^b\)Cohen’s d used for Effect Size(ES) and Standardized Response Mean(SRM); trivial <0.2, small =0.2 <0.5, moderate >=0.5<0.8, large >=0.8
Table 7.
Results of Post Hoc Analysis of Stationary Subtest by Time and Function\textsuperscript{a}, with Responsiveness Indices

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Function</th>
<th>Baseline Mean Percentage (SD)</th>
<th>Week 52 Mean Percentage (SD)</th>
<th>t value (adj. p)</th>
<th>ES\textsuperscript{b} (d)</th>
<th>SRM \textsuperscript{b}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stationary</td>
<td>1</td>
<td>60.6(1.7)</td>
<td>65.3(3.7)</td>
<td>1.46(.488)</td>
<td>1.63</td>
<td>1.91</td>
</tr>
<tr>
<td>Stationary</td>
<td>2</td>
<td>50.2(12.4)</td>
<td>60.0(3.1)</td>
<td>2.53(.105)</td>
<td>1.08</td>
<td>1.06</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Post Hoc analyses with Tukey-Kramer test for multiple comparisons

\textsuperscript{b} Cohen’s d used for Effect Size (ES) and Standardized Response Mean (SRM); trivial <0.2, small =0.2 <0.5, moderate >=0.5<0.8, large >=0.8

---

Table 8.
Results of Post Hoc Analysis of Gross Motor Composite by Time and Function\textsuperscript{a}, with Responsiveness Indices

<table>
<thead>
<tr>
<th>Composite</th>
<th>Function</th>
<th>Baseline Mean Percentage (SD)</th>
<th>Week 52 Mean Percentage (SD)</th>
<th>t value (adj. p)</th>
<th>ES\textsuperscript{b} (d)</th>
<th>SRM \textsuperscript{b}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross Motor</td>
<td>1</td>
<td>39.3(6.1)</td>
<td>55.4(8.5)</td>
<td>5.59(.0001)*</td>
<td>2.17</td>
<td>1.06</td>
</tr>
<tr>
<td>Gross Motor</td>
<td>2</td>
<td>30.4(13.1)</td>
<td>40.1(13.9)</td>
<td>3.38(.027)*</td>
<td>0.72</td>
<td>0.63</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Post Hoc analyses with Tukey-Kramer test for multiple comparisons

\textsuperscript{b} Cohen’s d used for Effect Size (ES) and Standardized Response Mean (SRM); trivial <0.2, small =0.2 <0.5, moderate >=0.5<0.8, large >=0.8
Table 9.
Example of PDMS-2\textsuperscript{a} Raw Scores at Baseline and Week 52 for Two Subjects

<table>
<thead>
<tr>
<th></th>
<th>Time</th>
<th>Subject A Functional Group One</th>
<th>Subject B Functional Group Two</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age at Baseline</strong></td>
<td></td>
<td>12m</td>
<td>15m</td>
</tr>
<tr>
<td><strong>Stationary</strong></td>
<td>Baseline</td>
<td>36</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Week 52</td>
<td>38</td>
<td>36</td>
</tr>
<tr>
<td><strong>OM\textsuperscript{b}</strong></td>
<td>Baseline</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Week 52</td>
<td>19</td>
<td>4</td>
</tr>
<tr>
<td><strong>Locomotion</strong></td>
<td>Baseline</td>
<td>68</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Week 52</td>
<td>113</td>
<td>44</td>
</tr>
</tbody>
</table>

\textsuperscript{a}PDMS-2: Peabody Developmental Motor Scales- 2\textsuperscript{nd} edition

\textsuperscript{b}OM-Object Manipulation
References


6. Services NCEI. New Eligibility Definition For the NC Infant Toddler Program http://www.ncei.org/ei/.


17. Middel B, Van Sanderen E. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. 2002.

Chapter III.
Concurrent Validity of the Peabody Developmental Motor Scales - Second Edition and the Pediatric Evaluation of Disability Inventory, the Pompe Pediatric Evaluation of Disability, and the Alberta Infant Motor Scale in Children with Pompe Disease

Introduction

Documentation of eligibility for therapy service and efficacy of intervention with standardized, norm-referenced assessment tools is increasingly being required for reimbursement. Eligibility for the North Carolina Infant Toddler Program currently requires children to have an established condition or developmental delay documented with scores of 2.0 standard deviations below the mean on standardized tests or 30% delay on an instrument which uses age equivalent test scores.\(^1\) *The Guide to Physical Therapy Practice* includes tests and measurements as an essential component in examination and evaluation, used to establish functional limitations, impairments and baseline information.\(^2\) Pediatric physical therapists are completing standardized assessments as part of their routine clinical practice, with 59% reporting that they use a standardized measure daily or weekly.\(^3\)

In a survey of pediatric physical therapists, the Gross Motor Function Measure 88 (GMFM), the Alberta Infant Motor Scale (AIMS) and the Peabody Developmental Motor Scales-2 (PDMS-2) were the three most commonly used pediatric motor outcome measurement tools.\(^3\) *The GMFM - 88 is a criterion
based assessment tool that does not provide a comparison to a normative sample. The PDMS-2 and the AIMS are both evaluative measures designed to assess motor development and provide normative percentile ranking and age equivalent scores.\textsuperscript{4} The PDMS-2 is appropriate for use from birth to 72 months of age and the AIMS from birth to 18 months of age.\textsuperscript{5}

Test selection, to determine eligibility for services, may vary regionally, and specific tests may be mandated for use at a county or state level. Therapists must have an awareness of test limitations and strengths and to what degree the scores are related to scores on other standardized tests. Lack of agreement between outcome measurement scores may change eligibility status for services between agencies and geographic regions and misinform efficacy intervention research. Concurrent validity is one method of evaluating the level of agreement that is present between two outcome measurements. The validity of tool use for a specific population is demonstrated if a strong relationship is established between the tool and a criterion reference tool.

PDMS-2 test developers evaluated concurrent validity between the PDMS-2 and the Mullen Scales of Early Learning (MSEL) in typically developing children. The MSEL provides a baseline measurement of cognition and motor development. The largest correlation coefficients were present between the MSEL and the PDMS-2 Gross Motor Quotient (\(r= 0.86\)) and between the MSEL and Locomotion and Object Manipulation subtests (both \(r= 0.90\)).\textsuperscript{4} The PDMS-2 Stationary subtest did not correlate significantly with the MSEL Gross Motor Scale.
Snyder et al (2008) examined the concurrent validity of the AIMS and the PDMS-2 gross motor subtests on infants at risk for motor delay from birth to 18 months of age.\textsuperscript{6} The Pearson product moment correlations varied from $r= .78$ to .97. The largest correlation coefficient of $r= .97$ was found for the Locomotion subtest in infants less than nine months of age.

Researchers have found that the Bayley Scales of Infant Development II (BSID-II) Psychomotor Index (PDI) and the PDMS-2 had the strongest concurrent agreement for the Locomotion subtest, and that inconsistency was present in the ability to determine significant delay with standard scores.\textsuperscript{7,8} The BSID-II – PDI is a discriminative motor scale that is used to identify developmental delay in children from one month to 3.5 years of age. Provost et al (2004) found that more than 75\% of the sample classified as significantly delayed on the BSID-II (PDI $\leq 69$) did not score in the classification of “very poor” (TMQ, $= 69$) on the PDMS-2.\textsuperscript{7} The PDMS-2 Locomotion subtest showed a high age equivalent correlation of $r= .97$ with the BSID-II PDI. The Locomotion subtest also showed the highest age equivalent agreement with the BSID PDI at 96\% within 3 months. The Stationary and Object Manipulation subtests had only a 90-95\% agreement within five months.

Connolly et al (2006) also examined PDMS-2 concurrent validity and found a low ($r= .30$) correlation between the standard scores on the PDMS-2 and the BSID-II in typically developing, 12 month old infants.\textsuperscript{8} No correlation was found between the age equivalent scores for the Stationary and Object Manipulation subtests, but a high and significant correlation ($r= .71$, $p< .05$) was
found between the age equivalent PDMS-2 Locomotion and the BSID-II motor scale.\(^8\) Lack of agreement between the two tests could represent variability in service eligibility.\(^7\) When tests are important for clinical decision-making and service eligibility, Provost recommends a very high level of correlation \((r= .95)\) between two tests that are used to qualify for services.\(^7\)

PDMS-2 developers reported lower mean item discrimination coefficients for the Stationary subtest than the Locomotion subtest.\(^4\) Item discrimination, as it relates to the PDMS-2, evaluates the subtest’s ability to discriminate between different levels of function. The Stationary subtest includes items to capture balance in sitting and kneeling, but no items are included to capture standing balance or transitions in and out of standing. The items progress from #19 maintaining balance in kneeling for five seconds, to #20 standing on one foot for three seconds. Each item in the PDMS-2 has a normative age equivalence score recorded in months. The age equivalence score for #19 is 13 months and for item #20 is 31-32 months. Therefore only one item is included with an age representation between 13 and 31 months. In contrast to the Stationary subtest, the AIMS provides five items that qualitatively analyze stationary standing.\(^5\) Large jumps in the activity level required to pass items decrease instrument precision. Item gaps may produce inaccurate age equivalent and standard score representation and low concurrent validity values when comparisons are made to other pediatric motor outcome measurement tools.

Mayrand et al (2009)\(^9\) examined the association between the PDMS-2 gross motor scale and the Pediatric Evaluation of Disability Inventory (PEDI)
Functional Skills Mobility domain in children with language impairment. The PEDI is a standardized instrument for children aged 6 months to 7.5 years that uses parent report to measure level of disability in self care, mobility and social function domains. The Functional Skills mobility domain contains items that are relevant for daily independence in mobility function such as toilet, chair, wheelchair, bed tub and car transfers; bed mobility; indoor and outdoor mobility, and stair ascent and descent. The study found a low, non-significant correlation \( r = .23 \) between the PEDI mobility and PDMS-2 GMQ when used with children with combined primary language impairment and mild motor impairment. The PDMS-2 is a motor outcome measure, whereas the PEDI focuses on the degree impairment impacts function in activities of daily living. Although motor capacity should reflect actual performance in ADL, the PEDI may not have adequate sensitivity to detect subtle motor impairment in children with a primary language impairment. Better agreement between the PDMS-2 and the PEDI might have been present in a population of children with more significant motor impairment.

Available research on the concurrent validity of the PDMS-2 focuses on typical development and children at risk for developmental delay. There is a paucity of evidence to illustrate the relationship between the PDMS-2 and other developmental outcome tools in children with chronic disease and moderate motor functional impairment.
Pompe Disease

Pompe disease is a rare lysosomal storage disease characterized by a deficiency of the enzyme acid alpha glucosidase (GAA).\textsuperscript{11} Lack of GAA causes accumulation of glycogen in cardiac, skeletal, and smooth muscle and central nervous system tissue, leading to progressive cardiomyopathy, respiratory compromise, generalized weakness and hypotonia.\textsuperscript{11} The natural history of infantile onset Pompe disease typically leads to death by 1 year of age.\textsuperscript{12} Infants treated with ERT, in the form of recombinant human acid $\alpha$ glucosidase (Myozyme) have been shown to have decreased ventilatory use and decreased risk of death.\textsuperscript{12} Developmental motor skill acquisition in response to Myozyme has been variable. Motor functional classification in response to Myozyme varies from independent ambulators, to functional sitters, to children that require caregiver assistance to complete all activities of daily living.\textsuperscript{12} Functional outcomes were measured in the pivotal Genzyme Corporation clinical trials for Myozyme (AGLUO 1602 and 1702) with concurrent administration of the PDMS-2, AIMS, PEDI and the Pompe Pediatric Evaluation of Disability Inventory (PPEDI).

The Pompe PEDI is a disease specific version of the PEDI that was developed to assess functional capabilities and performance in children with Pompe disease from 2 months of age through adolescence. Haley et al (2003) modified the original PEDI in 2003 in order to more accurately define the levels of physical functioning and disability that were present in children with Pompe disease.\textsuperscript{10} The PPEDI includes all of the items listed in the PEDI, as well as
additional items that are designed to reflect clinically relevant skills for children with Pompe disease. Items were added to the Functional Skills and Caregiver Mobility and Self Care scales to reflect the functional skills and deficits that were seen clinically in Pompe disease. Fifty-nine items were added to the mobility domain and 17 new items were added to the self-care domains. The items were added to raise the ceiling, lower the floor, and increase evaluation of assistive technology and adaptive equipment and to improve precision of scoring and potential sensitivity to change. Norm-based scoring was developed for these new items. Scoring algorithms for the PEDI were adjusted to reflect the new normative data for the Pompe PEDI.

Reliability and validity testing of the PPEDI was completed through telephone administration to 30 parents of children with Pompe disease. Subjects were identified on the Glycogen Storage Disease Network (GSD) list serve. Study participants were between six months and 22 years of age, primarily male (76%) and Caucasian (86%) and lived in the USA, Canada, Germany, Spain, and England, but the majority were from the USA. Only 10 of the 30 children had some ability to ambulate, and all children presented with functional skills below age expected normative values. Test re-test was done in a two week interval with a mobility domain intra class correlation coefficient of r= .98. Authors justified use of a separate assessment for mobility and self-care in Pompe disease due to the heterogeneous clinical functional presentation found in Pompe disease and the challenges of classifying the disease phenotypes.
Purpose

The purpose of this manuscript is to examine the concurrent validity of the PDMS-2 with the AIMS, the PEDI and the PPEDI in a sample of children with Pompe disease undergoing enzyme replacement therapy. The study involved secondary data from the Genzyme Corporation Pompe AGLUO 1602 and 1702 enzyme replacement therapy (ERT) clinical trials.

Use of this data set provides an opportunity to evaluate concurrently validity of the PDMS-2 in children with a significant level of motor impairment and variable levels of functional mobility. These data analyses may help to inform clinical decision making for Pompe disease and future clinical trial outcome design. It may also identify strengths and limitations of the PDMS-2 that have application to a larger range of developmental disabilities.

Innovative therapies are continually being developed to extend survival for children with rare diseases. Intervention efficacy needs to include not only survival and motor capacity outcomes but also indicators to measure level of disability in activities of daily living. Evaluation of the concurrent validity of the PDMS-2 and PEDI and PPEDI allows measurement of the relationship between gross motor skill performance in the clinic and function in the home and community environment for children with Pompe disease.

A secondary purpose is to evaluate the concurrent validity between the PEDI and the PPEDI. Numerous changes were made to the PEDI to make it more appropriate for use in Pompe disease. It has become widely used in the
Pompe literature and is recommended for use in the Pompe Registry, but limited research is available to support validity for use.

**Research Questions:**

1. In the AGLUO 1602 subjects, what is the relationship between the PDMS-2 gross motor subtest percentile and age equivalent scores and the AIMS percentile and age equivalent scores? Are the relationships similar for the baseline assessment at ≤7.2 months and the first assessment completed after 12 months of age?

2. In the AGLUO 1602 subjects, what is the relationship between the PDMS-2 GMQ percentile scores and the AIMS percentile scores?

3. In the AGLUO 1602 and 1702 subjects, what is the relationship between the PDMS-2 gross motor subtest percentage scores and the PPEDI Functional Skills Mobility scaled scores?

4. In the AGLUO 1602 and 1702 subjects, what is the relationship between the PPEDI Functional Skills Mobility scaled scores and the PEDI Functional Skills Mobility scaled scores?

5. In the AGLUO 1602 subjects, what is the relationship between the PPEDI Functional Skills Mobility scaled scores and the AIMS percentage scores?

Based on the review of literature, it was hypothesized that the strongest relationship involving the PDMS-2 would be present between the Locomotion
subtest and the AIMS, and that the strongest relationship would be present in the less than seven month age range.\textsuperscript{6,7,8} The PPEDI was hypothesized to have a stronger relationship with the PDMS-2 than the PEDI, due to the inclusion of additional items on the PPEDI to reflect clinically relevant skills for children with Pompe disease. A strong relationship was hypothesized between the PPEDI and PEDI Functional Skills Mobility dimensions. A stronger relationship between the PPEDI and the AIMS was hypothesized in the assessments at one year and older than in those for infants less than 7.2 months old.

**Participants**

The AGLUO 1602 clinical trial involved children from one month to 7.2 months of age at baseline. The AGLUO 1702 clinical trial involved children from six months to 42 months of age at baseline. Ten children from the AGLUO 1602 data had scores available to complete the AIMS analyses. The mean age for the first assessment period was 5.0 months (SD 2.2) and the mean age for the second assessment period was 16.0 months (SD 4.0). An additional seven children from AGLUO 1702 had data available to examine the relationship between the PDMS-2 and the PPEDI and PEDI. The AGLUO 1702 children had a mean age for the first assessment period (baseline) of 22.7 months (SD 11.5) and the second assessment period (week 52) of 34.8 months (SD 13.1). The overall gender distribution was 6 females and 11 males.
Procedure

The AGLUO 1602 and 1702 data were collected by physical therapists in the USA, Taiwan, Europe and Israel. The primary author of this manuscript collected USA clinical data and assisted with international reliability training. Reliability training included a review of the theoretical basis for each assessment, instruction on administration guidelines and data reporting, and establishing inter-rater reliability of 90% agreement for all of the tests. To minimize investigator bias, the assessors did not review previous scores prior to administration of a repeat assessment. The same therapist at each site administered all repeat assessments. All concurrent assessments were completed on the same day at each interval. Following administration of the assessment, score sheets were forwarded to central scoring and entered into a database.

PDMS-2

The PDMS-2 was administered according to the guidelines in the manual. Four subtests make up the PDMS-2 gross motor composite: Reflexes, Stationary, Locomotion, and Object Manipulation. The Reflex subtest is completed only for subjects less than 12 months of age, and Object Manipulation is used after 12 months of age. All items are scored on a three point scale, with 0 indicating that the criteria for successful performance were not met, 1 indicating the behavior is emerging, and 2 indicating successful performance of the item criteria. The PDMS-2 raw subtest scores were used to calculate standard percentile and age equivalent scores. PDMS-2 standard subtest scores and
motor quotients are calculated using normative age data. The gross motor subtest standard scores are used to calculate the overall Gross Motor Quotient (GMQ). Subtest raw scores were converted to percentage scores to complete the concurrent validity analysis with the PPEDI Functional Skills Mobility dimension scaled scores. The percentage score was calculated by dividing the obtained raw score by the maximum possible raw score and multiplying by 100.

AIMS

The AIMS was administered according to the guidelines provided in the manual. Administration involves observation of 58 items that measure weight bearing, posture and antigravity movement in supine, prone, sitting and standing positions. Typical play is observed with minimal handling. The items between the infant’s least and most mature item in each position create a “window” of current skills. Infants are given credit for all observed items in the “window” and for items that fall maturationally below the window. The total AIMS score is the sum of the four positional scores. The total AIMS score and the age at the time of assessment are used to determine the percentile ranking compared with a normative age matched sample.

Liao and Campbell described the AIMS as a valid instrument to measure motor ability in infants and to evaluate different positions in space. They examined the item structure of the AIMS and confirmed that the items found in each testing position were presented in order of difficulty. Low precision for differentiating between levels of function was found with age equivalent items.
greater than nine months of age. Authors stated that this was due to lack of items to measure change in performance after the child had achieved controlled lowering from standing to sitting. A systematic review of nine infant motor tests found that the AIMS demonstrated the strongest clinical utility and psychometric properties and was one of the best predictive measures for identification of atypical motor development.

**PEDI and PPEDI**

The PEDI is a standardized instrument for children aged 6 months to 7.5 years that uses parent report to measure level of disability in self care, mobility and social function domains. The PPEDI is a Pompe disease specific version of the PEDI. The PEDI and the PPEDI both measure Functional Skill and level of Caregiver Assistance in Self Care, Mobility and Social Function domains. Each item is scored 0 if the child is unable to complete and 1 if the child can successfully complete the item. The total raw, standard, and scaled score with standard errors can be calculated for each domain. The Functional Skills-Mobility domain contains items that are relevant for daily independence in mobility function such as toilet, chair, wheelchair, bed, tub and car transfers; bed mobility; indoor and outdoor mobility; and stair ascent and descent. The PEDI was administered following completion of the AIMS and the PDMS-2. Credit was given for items observed in the outcome measures. Parent report was used to complete the remaining items.
Data Analysis

The strength of the relationships between outcome measures was examined using correlation analysis with Pearson Product Moment correlation coefficients. The correlations were examined at baseline and at the first assessment period after one year of age for AGLUO1602 and at baseline and 52 weeks later for AGLUO1702. Each correlation analysis was interpreted using descriptive terms to define the strength of the relationship. Portney and Watkins criteria were used to judge the strength of the correlation coefficient, with 0.00-0.25 indicating little or no relationship, .26 to 0.50 indicating a fair degree of relationship, 0.51 to 0.75 indicating a moderate to good relationship, and 0.76-1.00 indicating a good to excellent relationship. The significance of the correlation coefficients was evaluated using p-values.

Research question #1, comparing PDMS-2 and AIMS scores, also included calculation of frequencies of agreements between the age equivalent scores on the AIMS and the PDMS-2 GMQ. Frequency of agreement was measured within one, two, three and four months.

Results

PDMS-2 and AIMS

The Pearson product-moment correlations between the PDMS-2 gross motor subtests and the AIMS age equivalent and percentile ranking scores are presented in Table 1. The age equivalent correlations indicate a good to excellent relationship (r = .75 -.95, significant p-values) in both age groups, for all
subtests except the Object Manipulation subtest ($r = .43$, p-value non-significant).

Only one of the percentile rank correlations was significant, the Locomotion subtest in the youngest age group ($r = .67$, p.033). A value of .67 indicates a moderate to good relationship between the Locomotion subtest and AIMS percentile rank scores.

Table 2 presents the mean AIMS and PDMS-2 gross motor subtest percentile ranks and age equivalent scores for each age group. The mean percentile rank in the $\leq 7.2$ month assessments was larger for the PDMS-2 gross motor composite (16.0) than the AIMS total score (11.6). The mean percentile rank in the $\geq$ one year assessments was larger for the AIMS total score (34.8) than the PDMS-2 gross motor composite (28.4). Table 3 shows the percent agreement for the age equivalent scores for the combined age groups.

**PDMS-2 and PPEDI/PEDI**

The Pearson Moment correlation coefficients between the PDMS-2 subtest percentage scores and the PPEDI Functional Skills scaled scores are presented in Table 4. For the children in the AGLUO 1602 clinical trial, significant correlations in the youngest age group were present only for the Locomotion ($r = .76$, p.01) and Stationary ($r = .83$, p.003) subtests. Both correlations fell in the good to excellent range. For the children who were one year of age and older, the values for the Locomotion ($r = .63$, p .051) and the Stationary ($r = .68$, p .029) subtests fell in the moderate to good range. Correlations for the Reflex subtest in the $\leq 7.2$ months age group and the Object Manipulation subjects in the $\geq$ one
year of age group were not significant. In the AGLUO 1702 subjects at week 52, the correlation coefficients increased for both the Locomotion and Stationary subtests and remained non-significant for the Object Manipulation subtest. A moderate relationship was demonstrated between the PPEDI Functional Skills Social dimension and the Stationary subtest \((r=0.60, p=0.012)\). The PPEDI Functional Skills Self Care dimension demonstrated a moderate to good relationship with the Locomotion subtest \((r=0.64, p=0.0057)\) and a good to excellent relationship with the Stationary subtest \((p=0.79, p=0.0002)\).

The correlations for the relationship between the PEDI and the PDMS-2 subtest percentage scores were Locomotion subtest \((r=0.90, p<0.0001)\), Stationary subtest \((r=0.85, p=0.0071)\), and Object Manipulation subtest \((r=0.21, p=0.5794)\). The values were very similar to the correlations between the PPEDI and the PDMS-2, Locomotion subtest \((r=0.76, p=0.0004)\), Stationary subtest \((r=0.83, p=0.0001)\), and Object Manipulation subtest \((r=0.33, p=0.1869)\).

**PPEDI and AIMS**

The Functional Skills Mobility dimension scaled scores and the AIMS had a moderate to good relationship \((r=0.70, p<0.0001)\) for children ≥one year of age, and a good to excellent relationship \((r=0.76, p<0.0001)\) for children ≤7.2 months of age.
**PPEDI and PEDI**

The PPEDI and PEDI Functional Skills Mobility dimension scaled scores had a moderate to good relationship \( r = 0.69, < 0.0001 \) for children \( \geq \) one year of age.

**Discussion**

**PDMS-2 and AIMS**

The hypothesis that the Locomotion subtest would have the strongest subtest relationship between the AIMS and the PDMS-2 was not supported by the age equivalent results. Although, similar to Synder et al (2008)\(^6\), the Locomotion values were slightly larger than the other subtests, all age equivalent values fell in the good to excellent range except for Object Manipulation.\(^6\) It is reasonable to expect a lower correlation value for the Object Manipulation subtest due to differences in item content between the tests. The Object Manipulation subtest contains ball throwing, catching and kicking skills and the AIMS contains no items with ball skills. The Stationary results are in contrast to the Provost et al (2004) study that found non-significant age equivalent correlations between the BSID-II and the PDMS-2 for the Stationary and Object Manipulation subtests.\(^7\) Provost was able to evaluate the PDMS-2 and BSID-II in a large age range, from three to 41 months. The range may have been more suitable for capturing possible standing limitations on the Stationary subtest.

The PDMS 2 developers suggest that caution should be used when interpreting test results with age equivalent scores.\(^4\) They recommend use of
standard and percentile scores when interpreting test results. This study found that the age equivalent correlations were higher than the percentile rank correlations. The percentile rank Locomotion subtest values were in the good and good to excellent range but the Stationary subtest had only a fair relationship in the youngest group (≤7.2 months) and a non-significant p-value for the older group (≥one year of age). Non-significant p-values were also found for the Reflex subtest in the youngest group and the Object Manipulation subtest for the older group. The percentile values from this research may support utility of the Locomotion subtest as a discriminate measure in children diagnosed with Pompe disease. Additional research with a larger sample size would be beneficial to determine with more precision the strength of the relationships in all of the subtests.

The relationship between the percentile rank scores for the PDMS-2 and the AIMS was lower in the ≤7.2 month than the ≥one year age group for the Stationary and Locomotion subtests. The AIMS also had a lower mean percentile rank than the PDMS-2 GMQ in the younger group. These findings might reflect differences in the item content of the AIMS and the PDMS-2 for the younger children. Infants with Pompe disease often have general hypotonia and proximal weakness. The AIMS contains items that qualitatively measure trunk strength in supine such as chin tuck with active abdominal muscles for hands to knees, and prone items that require not just a measurement of head elevation, but an active chin tuck with neck elongation and chest elevation. In contrast, the PDMS-2 contains items that measure head alignment relative to the trunk with examiner
facilitated movements or sitting postures, and prone is measured through head and upper trunk elevation as a timed task. The AIMS may contain more items than the PDMS-2 Stationary and Locomotion subtests to capture the specific weakness deficits that are seen in infants who present with Pompe disease.

The non-significant correlation between the AIMS and Stationary subtest percentile rank scores in the older age group may also be explained by the typical motor presentation of children with Pompe disease. Infants with Pompe disease often have more lower than upper extremity impairment and demonstrate delayed independent standing and ambulation. The Stationary subtest has an item progression from tall kneeling to standing on one foot independently. It lacks items to capture quality of movement in supported standing and emerging independent standing. The AIMS, in contrast, has numerous items to capture quality of standing such as rotating in supported standing, standing alone, and lowering from standing to the floor. PDMS-2 results for the Stationary subtest do not indicate whether the child's skills are closer to the tall kneeling item or the standing on one foot item. If a child was able to tall kneel, and stand at a support surface, but unable to achieve standing on one foot, the maximum item credit would be for tall kneeling. On the AIMS, the same child would receive credit for standing items, and the relationship between motor skill acquisition as measured by the two assessments would be reduced.

The correlation coefficient between the PDMS-2 GMQ and AIMS percentile rank for the combined age groups was statistically significant. In
previous investigations PDMS-2 individual subtest shortcomings may have been masked due to averaging of the three scores to obtain one test score.

**PPEDI**

A comprehensive evaluation of function in children with Pompe disease should not be limited to assessments that evaluate capacity to perform motor skills in the clinic environment. Function should also be considered within the context of environment. Activity limitations and participation restrictions within the community, home, and school environments should be considered. Correlation coefficients between the PPEDI Functional Skill Mobility dimension and the PDMS-2 Stationary and Locomotion subtests were in the good to excellent range. This demonstrates that motor capacity on the PDMS-2 reflects actual performance on the PPEDI. The hypothesis that the Locomotion subtest would have a stronger relationship with the PPEDI than the other subtests was not supported. All Locomotion and Stationary values fell into the moderate to good or good to excellent range. The Reflex subtest had a non-significant, fair relationship with the PPEDI. This finding may reflect differences in item content, because the PPEDI measures function in children with developmental disability, and the PDMS-2 Reflex subtest measures reflexes and righting reactions seen in typical development.

Correlation coefficients between the PPEDI Functional Skill Mobility dimension and the AIMS were also in the good range. This adds support for the capacity of the PPEDI to measure motor function in the home and community.
environment. A systematic review of nine infant motor tests found that the AIMS demonstrated the strongest clinical utility and psychometric properties and was found to be one of the best predictive measures for identification of atypical motor development.\textsuperscript{15}

The Object Manipulation subtest had a non-significant p-value at $\geq$ one year of age, but a good to excellent range value 52 weeks later. The PPEDI does contain ball skills similar to the Object Manipulation subtest, but the children at the younger ages may not have had the standing capacity to demonstrate the ball skills until they were older.

The PPEDI and PDMS-2 coefficient correlation values were much larger than those reported for the PEDI in children with language deficits and mild motor issues.\textsuperscript{9} One explanation for differing results is that children diagnosed with Pompe have much greater motor impairment than the mild motor issues present in the Mayrand et al population.\textsuperscript{9} Secondly, the inclusion of additional items on the PPEDI Functional Skills Mobility dimension, as compared to the original PEDI, may have increased the precision of the instrument for use in Pompe disease.

**Eligibility for Intervention Services.**

Eligibility for early intervention services requires providers to demonstrate developmental delay relative to a normative peer sample. Standards for determination of eligibility for early intervention services are variable between counties and states. Variability is also present in tool selection. The AIMS
developers did not provide a recommendation for definitive percentile rank interpretation to justify therapy referral for intervention. They recommended pairing a low percentile ranking with “ongoing monitoring, referral for diagnostic workup or intervention for motor delay” (page 49).5 Van Haastert et al (2006) used pediatrician developmental assessments to establish AIMS percentile values of normal, suspect, and abnormal in premature infants.18 The 10th percentile rank at four months of age produced a sensitivity of 77% in predicting abnormal movement at 18 months of age. If a 10th percentile rank was chosen as the most accurate level for determination of “abnormal movement” and qualification for early intervention services, five children in the present study would qualify using the PDMS-2 scores and ten would qualify with use of the AIMS scores. The PDMS-2 manual recommends interpretation of standard scores of < 79 on the GMQ as an indicator of “poor” gross motor movement.4 If PDMS-2 GMQ of 79 or less was used as the criterion for service eligibility in the present study, seven children would qualify for therapy services. If the AIMS and the PDMS-2 both used a percentile score cut off level < 25%, an equal number of twelve children would qualify for therapy services.

A lack of agreement between the AIMS and the PDMS-2 percentile scores and variable interpretive standards can create inequities in service eligibility between agencies and geographic regions. Therapists should be encouraged to concurrently administer more than one developmental motor outcome tests if possible. Documentation of outcome scores should be paired with detailed narratives and use of clinical judgments for referral to agencies. Multidisciplinary
education that includes physicians and parents should be completed to provide clarity on test interpretation.

**Limitations**

Direct application of these results is limited to a small population of children with Pompe disease. Concurrent validity was completed only on a small range of the PDMS-2, from 0 to 72 months administrative age. The sample size of this study was small, and non-significant p-values may have been due to inadequate statistical power. Additional research with a larger sample size is justified to determine the reproducibility of the results.

**Conclusion**

The study results demonstrate that, in children with Pompe disease, a stronger relationship was present between the PDMS-2 and AIMS age equivalent scores than the percentile scores. The relationship between the PDMS-2 and AIMS was strongest in the Locomotion subtest percentile scores.

Motor capacity on the PDMS-2 and the AIMS reflects actual performance in the home, school, and community environment as measured on the PPEDI in this sample of children with Pompe disease. PPEDI Functional skill Mobility dimension correlation coefficients between the PDMS-2 Stationary and Locomotion subtests and between the AIMS were in the moderate to good or good to excellent range. Item content differences may be present between the PPEDI and the Reflex and Object Manipulation subtests with administration to infants. The correlation coefficient results between the PPEDI and the Reflex and
Object Manipulation subtests in the <=7.2 month group indicated a non-significant, fair relationship.

A four-month range in age equivalent scores was necessary to achieve 100% agreement between the AIMS and the PDMS-2 gross motor subtests. Use of more than one developmental motor outcome measurement to determine eligibility for services is recommended.

Table 1.

Pearson Product Moment Correlation Coefficients (r) a for the Age Equivalent and Percentile Scores of the PDMS-2 Subtests and GMC b, and the AIMS c

<table>
<thead>
<tr>
<th>Score</th>
<th>Age Equivalent</th>
<th>Age Equivalent</th>
<th>Percentile Rank</th>
<th>Percentile Rank</th>
<th>Percentile Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation Age</td>
<td>≤7.2m</td>
<td>≥1yr</td>
<td>≤7.2m</td>
<td>≥1yr</td>
<td>Both Groups</td>
</tr>
<tr>
<td>Reflexes</td>
<td>.75 (.0117)*</td>
<td></td>
<td>.46 (.1774)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Locomotion</td>
<td>.91 (.0003)*</td>
<td>.97 (.0013)*</td>
<td>.67 (.0326)*</td>
<td>.87 (.1271)</td>
<td></td>
</tr>
<tr>
<td>Stationary</td>
<td>.86 (.0015)*</td>
<td>.96 (.0028)*</td>
<td>.36 (.3114)</td>
<td>.57 (.4324)</td>
<td></td>
</tr>
<tr>
<td>Object Manipulation</td>
<td></td>
<td></td>
<td>.43 (.4713)</td>
<td></td>
<td>.92 (.2531)</td>
</tr>
<tr>
<td>Gross Motor Composite</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.69 (.0032)*</td>
</tr>
</tbody>
</table>

a Pearson Product Moment Correlations displayed as r (p value)
c AIMS: Alberta Infant Motor Scale
Table 2.
Mean Age Equivalent and Percentile Rank Scores for the PDMS-2 GMC \(^a\) and the AIMS \(^b\)

<table>
<thead>
<tr>
<th>Score</th>
<th>Age Equivalent</th>
<th>Age Equivalent</th>
<th>Percentile Rank</th>
<th>Percentile Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDMS-2 GMC(^a)</td>
<td>2.1m</td>
<td>9.3m</td>
<td>16.0</td>
<td>28.4</td>
</tr>
<tr>
<td>AIMS(^b)</td>
<td>2.0m</td>
<td>10.4m</td>
<td>11.6</td>
<td>34.8</td>
</tr>
</tbody>
</table>

\(^a\) PDMS-2 GMC- Peabody Developmental Motor Scales Gross Motor Composite

\(^b\) AIMS Alberta Infant Motor Scale

\(^c\) m= Months

Table 3.
Percent Agreement in Age Equivalent Scores for the PDMS-2 GMC \(^a\) and AIMS \(^b\)

<table>
<thead>
<tr>
<th>Time</th>
<th>% Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 month</td>
<td>50.0</td>
</tr>
<tr>
<td>2 months</td>
<td>68.8</td>
</tr>
<tr>
<td>3 months</td>
<td>87.5</td>
</tr>
<tr>
<td>4 months</td>
<td>100.0</td>
</tr>
</tbody>
</table>

\(^a\) PDMS-2 GMC-Peabody Developmental Motor Scales Gross Motor Composite

\(^b\) AIMS-Alberta Infant Motor Scale
Table 4.

Pearson Product Moment Correlation Coefficients\(^a\) for the PDMS-2\(^b\) Subtest Percentage Scores and the PPEDI\(^c\) Functional Skills Dimension Scaled Scores

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AGLUO 1602 ≤7.2m</td>
<td>AGLUO 1602 ≥1yr</td>
<td>AGLUO 1602 &amp; 1702 ≥1yr</td>
<td>AGLUO 1702 Week 52</td>
<td>AGLUO 1602 ≥ 1yr &amp; 1702 Baseline</td>
<td>AGLUO 1602 ≥ 1yr &amp; 1702 Baseline</td>
</tr>
<tr>
<td>Reflexes</td>
<td>.49 (.155)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Locomotion</td>
<td>.76 (.010)*</td>
<td>.63 (.051)</td>
<td>.76 (.0004)*</td>
<td>.85 (.031)*</td>
<td>.26 (.318)</td>
<td>.64 (.0057)*</td>
</tr>
<tr>
<td>Stationary</td>
<td>.83 (.003)*</td>
<td>.68 (.029)*</td>
<td>.83 (&lt;.0001)*</td>
<td>.93 (.023)*</td>
<td>.60 (.012)*</td>
<td>.79 (.0002)*</td>
</tr>
<tr>
<td>OM(^d)</td>
<td></td>
<td>.39 (.268)</td>
<td>.34 (.1869)</td>
<td>.81 (.052)</td>
<td>-.10 (.710)</td>
<td>.04 (.8881)</td>
</tr>
<tr>
<td>Sample Size</td>
<td></td>
<td>10</td>
<td>10</td>
<td>17</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>Mean Age months (SD)</td>
<td></td>
<td>5m (2.23)</td>
<td>15.96m (3.96)</td>
<td>18.7m (8.6)</td>
<td>34.8m (13.1)</td>
<td>18.7m (8.6)</td>
</tr>
</tbody>
</table>

\(^a\) Pearson Product Moment Correlations displayed as \(r\) (\(p\) value)

\(^b\) PDMS-2- Peabody Developmental Motor Scales- 2\(^{nd}\) edition

\(^c\) PPEDI - Pompe Pediatric Evaluation of Disability Inventory

\(^d\) OM-Object Manipulation
References

1. Services NCEI. New Eligibility Definition For the NC Infant Toddler Program http://www.ncei.org/ei/.


glucosidase: major clinical benefits in infantile-onset Pompe disease. 

13. Wang HH, Liao HF, Hsieh CL. Reliability, sensitivity to change, and 
responsiveness of the peabody developmental motor scales-second 
1359.

14. Liao PJ, Campbell SK. Examination of the item structure of the Alberta 

15. Spittle AJ, Doyle LW, Boyd RN. A systematic review of the clinimetric 
properties of neuromotor assessments for preterm infants during the first 


alpha prolongs long-term survival of infants with Pompe disease. Pediatr 

18. Van Haastert IC, de Vries LS, Helders PJ, Jongmans MJ. Early gross 
motor development of preterm infants according to the Alberta Infant 
Chapter IV.
Synthesis

Summary of Results

The PDMS-2 is extensively used as an evaluative and discriminative measurement tool in clinical research, early intervention and rare disease clinics, and school systems. Golumb et al (2004) reviewed 517 articles on motor outcomes used in young children and found that the PDMS-2 was one of the four most commonly cited measurement tools.¹ Yet there is a paucity of validity research to support use with children with chronic diseases and moderate developmental disability, and guidelines are absent on how to interpret standard scores as a measurement of change in motor function. Secondary analyses of Genzyme Corporation Pompe disease AGLOU 1602 and 1702 data provided a unique opportunity to analyze responsiveness to change and concurrent validity of the PDMS-2 in children with a significant motor impairment.

Although the focus of this dissertation was to examine the PDMS-2 validity in children with Pompe disease, the identified limitations of use may be appropriate to consider with other diagnoses. The functional distribution of the sample was not dissimilar to what might typically be seen in pediatric clinic environments. Children with other diagnoses can present similar to functional group one with generalized weakness, delayed acquisition of independent ambulation and challenges with high-level balance and coordination skills, or
similar to functional group two with the ability to sit independently, required use of adaptive equipment to ambulate or stand, and primary community mobility with a wheelchair.

The results of this study support adequate responsiveness of the PDMS-2 in a heterogeneous group of children who present with Pompe disease. However, “responsive measures discriminate between trivial and substantial change within groups and between those groups.”² (page 3) When responsiveness to change was evaluated in two different functional presentations, the Locomotion subtest was able to discriminate change within both groups from baseline to week 52 and between groups at both time frames. The Stationary subtest was unable to discriminate between or within groups. The Object Manipulation subtest was able to discriminate change only within group one from baseline to week 52 and between functional presentations of group one and group two at week 52. The validity of a tool depends on the number of values on the instrument and the standard error of measurement of the instrument.³ The more items on the instrument, and the smaller the standard error of measurement, the greater the opportunity to detect change.³ Content validity defines the extent to which the items reflect a content domain and it is clear that no items are missing. The Stationary subtest may lack content validity and adequate responsiveness when level of function is considered due to an inadequate number of test items to quantify quality of bipedal standing and transitions in and out standing. Thirteen out of fourteen children had a week 52 raw score between 36 and 38. This represents a ceiling effect at either item #19,
maintaining balance in tall kneeling, or item #20, standing on one foot for three seconds.

An Object Manipulation subtest ceiling effect was also present for children who did not have the ability to balance in standing independently and complete ball skills. Children with a greater degree of developmental disability with an inability to maintain independent standing could not progress on the Object Manipulation items.

Acceptable responsiveness was present in the PDMS-2 percentage GMC scores when the functional groups were separated or combined. The GMC was calculated by summing the three gross motor subtests and then dividing by three to create an average score. A large change score in the Locomotion subtest could compensate for a small Stationary subtest change score and create an overall significant mean difference between the GMC at baseline and week 52. Although the testing time frame for this research was 52 weeks, clinical reevaluations for reauthorization of therapy services often need to be completed every three to six months. Over this shorter time frame, a smaller change in Locomotion subtest score may be present. A smaller Locomotion subtest change score, paired with the same lack of change on the other two subtests, may not produce a significantly different mean GMC change score. Wang et al (2006) did find acceptable GMC responsiveness with a three-month testing interval in children with cerebral palsy, but the research combined subjects with mild and severe cerebral palsy into one group and did not evaluate individual subtests.\textsuperscript{4}
Response indices values indicate that the Locomotion subtest and the GMC were more responsive to change in functional group one than two. The ES is calculated by dividing the mean change score by the pooled standard deviation of the baseline and week 52 scores. Functional group one had a larger mean change score and less variance in scores than functional group two, thus creating larger response indices. The Stationary subtest response indices were large in both functional groups and did not support lack of responsiveness to change. Effect size is affected by group variance, and the Stationary subtest had a very small score variance, thus creating large effect sizes. Therefore the ES and SRM should be interpreted with caution because the lack of items to capture standing function actually created a large ES.

Examination of concurrent validity of the AIMS and the PDMS-2, yielded results that demonstrate a stronger relationship between age equivalent scores than the percentile scores. Although the Locomotion values were slightly higher than the other subtests, all age equivalent values except Object Manipulation fell in the good to excellent range. The PDMS-2 developers suggest that caution should be used when interpreting age equivalence scores and recommend use of percentile rank scores. The percentile rank Locomotion subtest values were in the good, to good to excellent range but the Stationary subtest had only a fair relationship in the youngest group and a non-significant correlation coefficient for the older group. This lends some support to research that has been completed in children with mild developmental delay that suggests that the Locomotion subtest
may have the best subtest discriminative utility. Additional research with a larger sample size is necessary to determine the reproducibility of the results.

Eligibility for early intervention services requires providers to demonstrate developmental delay relative to a normative peer sample. Variability in tool selection and standards for determination of eligibility for therapy services are present between evaluation centers. The AIMS developers do not provide a recommendation for definitive percentile rank interpretation for referral to therapy. Van Haastart et al (2006) used pediatrician developmental assessments to establish AIMS percentile values of normal, suspect and abnormal in premature infants. The 10th percentile rank at four months of age produced a sensitivity of 77% in predicting abnormal movement at eighteen months of age. If a 10th percentile rank was chosen as the most accurate level for determination of “abnormal movement”, and qualification for early intervention services, five children would qualify using the PDMS-2 scores and ten would qualify with use of the AIMS scores. A four-month range in age equivalent scores was necessary to achieve 100% agreement between the AIMS and the PDMS-2 in this research. A lack of agreement between the AIMS and the PDMS-2 percentile and age equivalent scores and variable interpretive standards can create inequities in service eligibility between agencies and geographic regions. Therapists should be encouraged to concurrently administer more than one developmental motor outcome test when possible. Documentation of outcome scores should be paired with detailed narratives and use of clinical judgment.
Inclusion of the PPEDI data in this dissertation provided an opportunity to examine a tool that is very commonly used in the Pompe literature and is recommended for use in the Pompe Registry.\textsuperscript{7} A comprehensive evaluation should include not just capacity to perform motor tasks in the clinic environment. It also should identify activity limitations and participation restrictions in the home, school, and community environments. The only validation research available for the PPEDI was completed in the initial test development. Given the PDMS-2 shortcomings that have been identified in this research, it was beneficial to also compare the PPEDI to the AIMS. A systematic review of nine infant motor tests found that the AIMS demonstrated the strongest clinical utility and psychometric properties and was found to be one of the most predictive measures for identification of atypical motor development.\textsuperscript{8} The PPEDI Functional Skill Mobility dimension was significantly correlated with the PDMS-2 Stationary and Locomotion subtests and with the AIMS, with correlation coefficients in the good to excellent range. This demonstrates that motor capacity on the PDMS-2 and the AIMS reflects actual performance on the PPEDI.

**Application to Clinical Practice and Research**

The results from this research will inform future Pompe disease protocol development. The clinical development team at Genzyme could not use the PDMS-2 data for FDA drug approval because they had no idea how to interpret the standard scores in efficacy research. They used raw AIMS scores beyond the age of recommended use to demonstrate motor skill acquisition in response to ERT. The published manuscripts from the clinical trials documented motor skill
acquisition without inclusion of any specific motor data. In view of the results from
the present study, raw scores could be converted to percentage scores to
demonstrate efficacy on the basis of changes in the GMC score and the
Locomotion subtest score over the 52-week period.

Conversion of the PDMS-2 subtest raw scores into percentage scores
may be beneficial for interpretation as an evaluative measure in other
populations. Norm referenced standard scores are appropriate for use as a
discriminative tool, but very challenging to use in evaluating change. Accurate
standard score interpretation by health care and school professionals, insurance
agencies, and parents is essential to guide decision making and plan of care
development. Normative mean standard scores are variable on pediatric motor
tests, and interpretation of change is confusing. Professionals and parents may
be unaware that the same standard score on two consecutive testing periods
actually represents motor skill acquisition. The raw scores in this research were
easily converted into percentage scores. PDMS-2 percentage score use as a
measurement of change in performance would be very similar to the percentage
scores on the GMFM or the Scaled scores on the PEDI.  

This research highlighted the importance of functional classification of
subjects in Pompe disease efficacy research. Ongoing development of more
detailed functional classification system for use in Pompe disease is
recommended. The PDMS-2 may be most appropriately used in children with
mild developmental delay, who have the capacity or potential to; balance in
standing to complete ball skills and acquire higher standing skills like standing on
one foot. Use of the Locomotion subtest as a discriminative measure to define function relative to a normative sample may also be appropriate. A wide heterogeneity is present in the functional presentation of children with Pompe disease, and functional course is not always predictable. If a study sample primarily involved children with significant motor impairment, adequate subtest responsiveness to change many not be present in the Stationary and Object Manipulation subtests.

Multidisciplinary education that includes physicians, parents, and insurance providers should be completed to provide clarity on test interpretation. Therapists must be educated in tool limitations in order to explain why a clinically significant change was observed but not represented on the standardized test scores or why variability exists between interagency evaluations.

**Limitations and Future Research**

The primary limitation is that the results from both manuscripts are directly applicable only to the small population of children with Pompe disease. The concurrent validity examination is also applicable only to a partial range (0=18months) of the age span in which the PDMS-2 use is recommended for use (0-72months). The AIMS is appropriate for use only until 18 months of age. Additional studies on concurrent validity would be beneficial for the age group beyond 18 months of age. The BSID-III contains a gross motor subtest and is normed for use from birth to 64 months of age. It contains standard scores and
percentile scores that could be concurrently examined with the PDMS-2 in future research.

An additional limitation is that the 52-week interval between testing periods is longer than what is typically required for insurance authorization. It is, however, typical of the efficacy clinical trials for ERT.

It would be very interesting to look at a Rasch analysis of item fit in the Stationary and Locomotion subtests independently and then combined to create one motor subtest. Fluid administration in the clinic environment requires therapists to combine the subtests. The Object Manipulation subtest is distinct from the other two subtests because it just evaluates ball skills.
References


2. Middel B, Van Sanderen E. *Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research*. 2002.


Appendix One: Literature Review

PDMS-2

The PDMS-2 is a standardized, norm-referenced test that contains four subtests that make up the gross motor composite: reflexes, stationary, locomotion and object manipulation; and two subtests that make up the fine motor composite: grasping and visual-motor integration. The total motor composite includes items from all of the subtests. All items are scored on a 3 point scale, with 0 indicating that the criteria for successful performance were not met, 1 indicating the behavior is emerging and a 2 indicating successful performance of the item criteria. The subtest raw scores can be used to calculate standard, percentile, z, and age equivalent scores. The subtest standard scores can be used to calculate gross, fine, and total motor composite quotient, percentile, and age equivalent scores. Gross, fine, and total motor quotient scores have a mean of 100 with a standard deviation of 15. The test developers recommend use of the motor quotient and percentile scores for interpretation of test results.

Inter-rater reliability of the PDMS-2 has been documented in the research for typically developing children (ICC of r= .96-.99), children with cerebral palsy (ICC of r= .88-1.0), and children with Hurler’s Syndrome (ICC of r= .74-.98) for the GMQ, raw scores and age equivalent scores. Reliability defines the accuracy or repeatability of the test when administered on different occasions or by different test administrators. Reliability is often misinterpreted by physical therapists as the only key component in choosing a test. Increased reliability
does increase the probability of detecting a difference between two groups, but it will not tell us if the variable is measuring what we want it to. A test also can be reliable but lack validity in terms of measuring important key constructs.

The *Taxonomy of the Psychomotor Domain* by Harrow (1972) was the foundational construct for the PDMS-2. The taxonomy assumes that development follows a hierarchical sequence of skill acquisition that begins with reflexive behavior and progresses through basic fundamental movements. The PDMS-2 test administration requires basal level item entry, with successful credit for three consecutive 2’s, and ceiling level discontinuation, with three consecutive 0’s. Children are credited for all items below the basal level, but not items above the ceiling level. This is in contrast to the Alberta Infant Motor Scale (AIMS), another commonly used norm-referenced tool. The AIMS is based on the dynamic systems model, and children are given credit for all items observed on the measurement tool. The hierarchical design of the PDMS-2, with lack of credit for skills above the ceiling level, may limit validity due to an inability to capture motor functional ability in children who do not demonstrate a predictable rate or sequence of skill acquisition.

An outcome measure is valid if it measures what it is designed to measure and accurately reflects the clinical findings. Validity has also been defined as the evaluative summary of the evidence for use and the potential consequences of score interpretation. The validity of a tool is not an absolute value, because it varies according to the purpose of use and the population of subjects tested.

The PDMS-2 continues to expand in its clinical and research application.
Measurement of the PDMS-2 test validity should be an ongoing process to define appropriate use with a conclusive body of research. Construct, content, and criterion validity, as well as responsiveness, are key components that will be reviewed in the PDMS-2 manual and literature.

Construct validity indicates the degree to which a measurement reflects an operational definition.\(^7\) The construct validity of an instrument may evolve over time with changes in the understanding of the construct of interest. Based on the *Taxonomy of the Psychomotor Domain* the PDMS-2 developers defined nonlocomotor movements (stationary subtest) as “movements of the limbs and trunk, bending, extending stooping and twisting.”\(^1\) Conceptually for physical therapists, “nonlocomotor” or stationary skills may be considered as static balance or the ability to maintain center of mass within limits of stability without having to use a protective or equilibrium strategy. Test administrators may equate the term stationary more with the ability to maintain balance in sitting and standing than “movements of the limbs and trunk, bending, extending stooping and twisting”. There are items on the Stationary subtest that measure sitting and kneeling balance, but there are no items on the subtest that capture standing balance abilities. Construct validity for the PDMS-2 was reported in the test manual through a confirmatory factor analysis of the subtests and motor composites. Confirmatory factor loading for the 12-72 month gross motor composite was .79 for the Locomotion subtest and only .52 for Stationary and .40 for Object Manipulation subtests.\(^1\) Confirmatory factor analysis implies that researchers have specified which items are related to each subtest, and then
factor loading confirms or refutes representation of that factor (subtest) by the variable (items). Garson (2010) recommended loading of .70 or higher in confirmatory factor analysis to confirm that variables are reflected by the factor. This brings into question whether the stationary subtest captures the theoretical construct of its’ current use and interpretation in the clinic setting. Additional research with exploratory factor analysis would be necessary to evaluate whether several standing items currently included in the Locomotion subtest might factor favorably on the stationary subtest.

Content validity defines the extent to which the items reflect a content domain and it is clear that no items are missing. A measure has content validity when the items are randomly chosen from a list of items that could represent the variable. The PDMS-2 test developers report content validity with mean item discrimination coefficients in each subscale. In the Stationary subtest, the lowest discrimination coefficient, $r = .41$, was found in the 36-47 months age group. In the Object Manipulation subtest, the lowest correlation coefficient of $r = .36$ was reported in the 12-23 month range. In the Stationary subtest, an item progression is present from item #19 maintaining balance in kneeling for 5 seconds to item # 20 standing on one foot for 3 seconds, with no items in static standing. Only one item is included with an age representation between 14 and 42 months. In contrast, the AIMS provides five items that qualitatively analyze stationary standing. In the Object Manipulation subtest, item administration progresses from item #3 rolling a ball in sitting to item #4 independently maintaining standing and flinging a ball 5 feet. On the Stationary and Object
Manipulation subscale, item gaps may decrease the instrument’s ability to measure the construct of interest and the tool’s responsiveness to change.

Responsiveness of a tool can be defined as the ability to measure clinically important change over time. It depends on the number of items and the standard error of measurement of the instrument. The more items included on the instrument, and the smaller the standard error of measurement, the greater the opportunity to detect change. Wang et al (2006) evaluated the responsiveness to change of the PDMS 2 in children with cerebral palsy over a three-month period using paired t-tests, effect size, standardized response mean and the Guyatt Responsiveness Index. Subtest raw scores were used to calculate percentage scores for the subtests. Subtest percentage scores were then used to calculate gross, fine and total motor composite percentage scores. Paired t-tests demonstrated that a statistically significant change (p<.001) was present between the gross, fine and total motor percentage scores over 3 months. The effect size for the gross motor subtest percentage score was labeled as small at 0.2 according to Cohen standards. The standardized response mean for the gross motor composite was labeled as trivial at 0.9. Information was not included on individual subtests within the fine or gross motor composites. A PDMS-2 review of the literature did not provide any additional research on the responsiveness of the PDMS-2 gross motor subtests or composite. Responsiveness research is applicable only to the sample population tested.
Criterion validity is the degree to which the instrument reflects or is related to scores obtained on a well-established instrument or “gold standard”. Concurrent validity is a method of criterion validity that evaluates the agreement between the results of two instruments administered within the same time period. Correlation coefficients from $r= .80-.91$ were found between PDMS-2 standard subtest scores and the Mullen Scales of Early Learning (MSEL) standard gross and fine motor scale in the normative sample. The authors stated that the coefficients support equivalency between the tests, but no coefficient was provided between the MSEL gross motor value and the PDMS-2 Stationary subtest value, and the manual labeled the coefficient between the MSEL and the Stationary subtest as non-significant.

The PDMS-2 correlated well ($r=0.76$) with the Movement Assessment Battery for Children (MABC) in a sample of children with developmental coordination disorder between 4 and 6 years of age. The MABC contains 8 items divided into manipulative skills, ball skills and balance skills. The MABC was designed to evaluate children with mild to moderate motor impairment, and incorporates a standardized performance test and a criterion referenced observational checklist. The study excluded all subjects with a severe motor impairment such as cerebral palsy.

Synder et al (2008) examined the concurrent validity of the AIMS and the PDMS-2 gross motor subtests in infants who were at risk for motor delay, from birth to eighteen months of age. The Pearson product moment correlations varied from $r= .78$ to 97. The largest most significant correlation coefficient of $r=.$
97 was found for the locomotion subtest in infants less than nine months of age. The emphasis in the AIMS is on qualitative evaluation of movement in supine, prone, sitting and standing, and includes transitional movements in each of the positions. The locomotion subtest of the PDMS-2 includes transitional movement items. Inclusion of transitional items may provide an explanation for the higher correlation of the AIMS with the Locomotion subtest than with the Stationary subtest of the PDMS 2.

Provost et al evaluated the concurrent validity between the Bayley Scales of Motor Development- second edition (BSID-II) motor scale and the PDMS -2 in children who were referred for developmental evaluations.\textsuperscript{11} The BSID-II is a widely used discriminative and evaluative tool for children between 0 and 42 months of age. The BSID-II and the PDMS-2 are often used to qualify children for eligibility for physical therapy services in early intervention. A correlation of \( r=0.75 \) between the BSID-II Psychomotor Index (PDI) and the PDMS-2 Gross Motor Index was demonstrated. However, the tests were inconsistent in their ability to determine significant delay with standard scores. More than 75% of the children who were classified as significantly delayed on the BSID-II (PDI <=69) did not score in the classification of “very poor” (TMQ =69) on the PDMS-2.\textsuperscript{11} The PDMS-2 Locomotion subtest showed a high age equivalent correlation of \( r=.97 \) with the BSID-II PDI. The Locomotion subtest also showed the highest age equivalent agreement with the BSID PDI at 96% within three months. The Stationary and Object manipulation subtests had only a 90-95% agreement within five months. A 5-month delay in a 20-month-old child equates to a 25%
delay in gross motor skills and supports eligibility for intervention services. Lack of agreement between the two tests could represent variability in service eligibility. When tests are important for clinical decision making and service eligibility, Provost recommends a very high level of correlation \((r = .95)\).\(^{11}\)

Concurrent validity between the PDMS-2 and the Bayley Scales of Infant Development-2 (BSID-II) also has been evaluated in typically developing, 12 month old infants.\(^{12}\) The standard scores on the PDMS-2 GMQ and the BSID-II Motor Scale were not significantly correlated \((r = .30)\). No correlation was found between the age equivalent scores for the Stationary and Object Manipulation subtests, but a high and significant correlation \((r = .71 \ p < .05)\) between the age equivalent PDMS-2 Locomotion and the BSID-II motor scale was found.

Literature is available to support concurrent validity of the PDMS-2 in typically developing children or children with mild developmental delay, but no research is available that supports concurrent validity in children with a chronic disease or moderate motor impairment.\(^{1, \, 5}\)

**PDMS-2 Summary**

The PDMS-2 is a widely used discriminative and evaluative measurement tool that covers the age range from birth to 72 months. Evidence is lacking to support the validity of using the PDMS-2 as an evaluative measure for infants and children with a moderate or severe motor impairment. Construct, content, and concurrent validity studies document poorer psychometric properties for the stationary than the locomotion subtest. The Stationary subtest age equivalence
item presentation gap from 13 months to 31-32 months, with lack of items to measure standing balance, may impact responsiveness to change. Large jumps in activity level required to pass items can decrease instrument precision. True functional level cannot be determined because it is unclear if ability is close to the passed item or close to the failed item. Lack of responsiveness to change may be the most apparent for the child who has activity limitations in independent standing balance.

**Pompe Disease and the AGLU0 1602 and 1702 Clinical Trials**

Pompe disease is a rare lysosomal storage disease characterized by a deficiency of the enzyme acid alpha glucosidase. (GAA) Lack of GAA causes accumulation of glycogen in cardiac, skeletal, and smooth muscle and central nervous system tissue, leading to progressive cardiomyopathy, respiratory compromise, generalized weakness and hypotonia. The natural history of infantile onset Pompe disease typically leads to death by 1 year of age.

Muscle function in Pompe disease may be reduced by a combination of muscular and neurological impairments. Glycogen accumulation has been found not only in the lysosomes, but also in the anterior horn cells, motor nuclei of the brain stem, and spinal ganglia. De Ruisseau et al found greater glycogen accumulation in the spinal cord than the brain, and noted absent or decreased deep tendon reflexes in individuals with the disease. Electromyography and nerve conduction studies have confirmed the presence of peripheral polyneuropathy in a two-year-old child with infantile Pompe disease.
Numerous articles document greater lower extremity than upper extremity motor impairment in Pompe disease. The studies are generally based on observation of decreased or absent ambulatory skills with preservation of upper extremity functional skills. No literature has used a standardized pediatric assessment to compare gross motor versus fine motor skill acquisition.

The Genzyme Corporation Pompe disease ERT clinical trials, AGLUO-1602 and AGLUO-1702, involved administration of ERT with recombinant human alpha glucosidase (rhGAA- Myozyme) to infants and children from birth to 36 months of age at baseline. The study’s primary objective was to evaluate the safety profile of rhGAA as determined by the proportion of patients alive and ventilator free over the course of the treatment. A secondary efficacy endpoint was the effect of treatment on motor development from baseline, as measured by the AIMS and/ or the PDMS-2. The Pediatric Evaluation of Disability Inventory (PEDI) and the Pompe Pediatric Disability Inventory (PPEDI) measured change in disability index from baseline. A control group was not used. Previously, an epidemiologic study of the natural history of Pompe disease was completed to provide a historical control for the AGLUO-1602 and AGLO-1702 clinical trials.

Many factors contribute to ERT response in infants and children with Pompe disease, including age, disease severity, age of initiation of ERT, GAA activity as measured by cross-reactive immunologic material (CRIM) and anti-rhGAA antibody titer levels. The infants who began receiving ERT before 6 months of age, but were unable to form a native enzyme GAA, responded poorly
to treatment in the trials, with no significant motor changes on the AIMS or PPEDI. Patients with a higher baseline median muscle GAA at baseline, an age of less than 12 months, and better baseline motor scores had the best response to ERT.

The AGLUO-1602 summary manuscript by Kishnani et al (2007)\textsuperscript{15} reported that 100% of the infants survived to 18 months of age. Compared to the untreated historical control group, the risk of death was reduced by 99%, and use of ventilatory assistance was reduced by 88%. Thirteen of 18 children from AGLUO-1602 demonstrated motor and functional changes on the AIMS and PPEDI. At week 52, 7 children demonstrated the ability to walk independently, 3 children could pull to stand independently and walk with one hand held, and 3 children could sit and roll independently but were not able to demonstrate weight bearing in standing.\textsuperscript{15} The remaining five subjects did not demonstrate clinically meaningful change on the AIMS or PPEDI. The children with the largest change scores on the AIMS also demonstrated the most substantial functional gains on all three PPEDI domains (self-care, mobility and socialization).\textsuperscript{15}

In the AGLUO-1702 clinical trial 16 of 21 children were alive at the end of 52 weeks.\textsuperscript{21} Five patients died before week 28 due to cardiac or respiratory failure. At baseline, 5 children were invasively ventilated and 2 were noninvasively ventilated (via mask). Of the 5 children who were invasively ventilated, 3 remained ventilated 24 hours a day, 1 reduced ventilation to 12 hours a day and 1 child died. Nicolino et al (2009)\textsuperscript{21} documented an age equivalent motor change score in 13 of 21 children on the AIMS or PDMS-2 but
no scores or data analyses were included in the manuscript. The researchers reported that five children were able to walk independently and eight patients could sit independently. The remaining eight patients made no significant motor development gains from baseline.

There is not an established level of clinically meaningful change for the AIMS, the PEDI or the PPEDI. The test developers for the PEDI recommend use of the standard error in interpretation. Greater certainly is present in equating a change score to improved functional performance if it exceeds 2x the standard error. Fourteen of 18 children in the AGLUO1602 trial had PPEDI scaled change scores that exceeded 2x the standard error. Only one child out of 18 had a decline in skill on the AIMS but an improvement on the PPEDI.

Ten of 16 children in the AGLUO 1702 trial had a PPEDI scaled score change that exceeded 2x the standard error. Eight of the 10 children also demonstrated new item acquisition on the AIMS. The AIMS may not be as useful in clinical trials for Pompe disease with the child who is lower functioning because it measures quality of movement along a typical development continuum. The requirements for antigravity movement limit item credit for many of the children who are more severely impaired with Pompe disease. Limited item credit may also be present for the child with non-classical presentation and compensatory movement patterns who does not follow a typical developmental pattern. AIMS responsiveness could also have been limited for children who were ventilator-dependent at baseline and have no tolerance for prone positioning. The AIMS requires observation of gross motor skills in supine prone,
sitting and standing. Elimination of ¼ of positional items could limit tool
responsiveness.

**Administration Procedure for the AIMS, PEDI, PPEDI and PDMS-2**

The PDMS-2 was concurrently administered with the AIMS, PEDI and
PPEDI to 18 subjects under the age of 6 months of age (AGLUO -1602) and 21
children between 6 months and 36 months of age (AGLUO -1702). Outcome
measurement testing was completed at baseline, week 12, 26, 38, and 52. The
PDMS -2 was administered and recorded, but never analyzed as a primary
outcome measure in the literature.

The primary author of this proposal collected the USA clinical data and
participated in international reliability training that included physical therapists
from Taiwan, Europe and Israel. Each physical therapist or physiotherapist who
participated in the clinical trial received training regarding the specific outcome
measurement tools being utilized and established inter-rater reliability. Training
included a review of the theoretical basis for each assessment, preparation for
and administration of each assessment, and logistics for reporting results. To
minimize inter-rater variability in the current study, the same therapist
administered the assessments to the same subjects. The AIMS and the PDMS-2
employ objective definitions for item credit and a standardized scoring system.
The PEDI and PPEDI manuals also contain standard scoring guidelines for item
credit. The parent or primary caregiver was interviewed for the PEDI and PPEDI.

In addition, professional observation of each child’s functional behavior
was used. The same parent/caregiver was interviewed on subsequent assessments. Video assessments of a child with Pompe disease and a typically developing child were used to establish inter-rater reliability of 90% agreement for all of the tests. To minimize investigator bias, the assessors did not review previous scores prior to administration of a repeat assessment. Following administration of the assessment the score sheet was forwarded to central scoring for tabulation of total raw, standard, scaled, age equivalent and percentile scores for the individual tests.

**Additional Instruments**

**AIMS**

The AIMS is an observation tool for examination of postural control in infants. The intended purposes are to identify infants with developmental delay, to address the rate of motor development with repeated testing, and to identify infants with abnormal patterns of movement. Use is recommended for infants from birth through acquisition of independent walking. Administration involves observation of 58 items that measure weight bearing, posture and antigravity movement in supine, prone, sitting and standing. The examiner observes typical play with minimal handling for the “window of function,” and then infants are given credit for the items in the window and those that fall maturationally below the window. To determine the total AIMS score the sum of the four positional scores is calculated. The total AIMS score and the age at the time of assessment are used to determine the percentile ranking compared with a normative age matched sample. Numerous articles look at the reliability and
validity of the AIMS, but recommendations have not been made for the
determination of clinically important change.\textsuperscript{13}

Liao and Campbell examined the item structure of the AIMS and
confirmed that the items found in each testing position were presented in order of
level of difficulty. \textsuperscript{13} Low precision for differentiating between levels of function
was found for age equivalent items greater than 9 months of age. This was
primarily due to lack of items to measure change in performance after the child
had achieved controlled lowering from standing to sitting. Overall, the
researchers characterized the AIMS as a valid instrument to measure motor
ability in infants and to evaluate different positions in space. They recommended
use of an additional tool like the PDMS-2 to measure motor development once a
child has demonstrated the ability to transition in and out of standing
independently.

**PEDI**

The PEDI is a standardized instrument for children aged 6 months to 7.5
years that uses parent report to measure level of disability in self care, mobility
and social function domains. \textsuperscript{23} Each of the three domains is measured as a
Functional Skill with respect to the level of Caregiver Assistance required. Each
item is scored 0 if the child is unable to complete and 1 if the child can
successfully compete the item. The total raw, standard, and scaled score with
standard errors can be calculated for each domain.
The functional skills-mobility domain contains 59 items that are relevant for daily independence in mobility function such as toilet, chair, wheelchair, bed, tub, and car transfers; bed mobility; indoor and outdoor mobility; and stair ascent and descent. Although the PEDI focuses on level of independence in mobility and ADL, research supports a high intra-class correlation coefficient for concurrent validity with the GMFM, an evaluative tool designed to measure gross motor function.

Mayrand et al evaluated the association between the PDMS-2 gross motor scale and the Pediatric Evaluation of Disability Inventory (PEDI) functional skills mobility domain in children with language impairment. The study found a low, non-significant correlation (r=. 23), between the PEDI functional skills mobility domain and PDMS-2 gross motor scale in children with language impairment. The PEDI was designed for use with children with physical disabilities or combined physical and cognitive disability. Use may not be as appropriate for children with a primary language impairment and mild motor impairment.

**PPEDI**

The Pompe PEDI is a disease specific version of the PEDI that was developed to assess functional capabilities and performance in children with Pompe disease from 2 months of age through adolescence. Haley et al (2003) modified the original PEDI in order to more accurately define the level of physical functioning and disability that was present in children with Pompe disease.
The PPEDI includes all of the items listed in the PEDI as well as additional items that are designed to reflect clinically relevant skills for children with Pompe disease. Items were added to the Functional Skills and Caregiver mobility and self-care scales to reflect the functional skills and deficits seen clinically in Pompe disease. Fifty-nine items were added to the mobility domain and 17 new items were added to the self-care domains. The items were added to raise the ceiling, lower the floor, and increase evaluation of assistive technology and adaptive equipment and to improve precision of scoring and potential sensitivity to change. Norm based scoring was developed for these new items. Scoring algorithms for the PEDI were adjusted to reflect the new normative data for the Pompe PEDI.

The PPEDI was administered by telephone interview to 30 parents of children with Pompe disease. Subjects were identified on the Glycogen Storage Disease Network (GSD) list serve. Study participants were between 6 months and 22 years of age, primarily male (76%) and Caucasian (86%) and lived in the USA, Canada, Germany, Spain, and England, with the majority from the USA. In the sample, only 10 of 30 children had some ability to ambulate, and all children presented with functional skills below age expected normative values. Test re-test was done in a two week interval with a mobility domain intra class correlation coefficient of $r= .98$. Authors justified use of a separate assessment for mobility and self-care in Pompe disease due to the heterogeneous clinical functional
presentation found in Pompe disease and the challenges of classifying the disease phenotypes.²³

**Summary**

The AGLUO 1602 and 1702 data provide a unique opportunity to use the PDMS-2 to evaluate, in greater detail, the functional presentation of infants and children with Pompe disease. Motor gains in response to ERT have been documented in the literature without inclusion of a defined level of meaningful change or outcome measurements comparing upper extremity and lower extremity function. Additional functional outcome data may help to define the impairment distribution between upper and lower extremity.

The validity of a tool is not an absolute value because it varies according to the purpose of use, and the population of subjects tested.⁵ Validity of use in children with Pompe disease has not been established with the PDMS-2. These data provide an opportunity to evaluate two key components of validity, responsiveness and concurrent validity.

The responsiveness of the PDMS-2 may be variable according to level of functional mobility. The stationary subtest item presentation gap may impact responsiveness to change as an evaluative measure. Large jumps in activity level required to pass items decreases the precision of the instrument.¹³ Lack of responsiveness to change may be most apparent for children who present with activity limitations in independent standing balance. The large age range of the children and a heterogeneous level of functional mobility in the sample provide
an opportunity to measure responsiveness of the PDMS-2 in children with Pompe divided by level of functional mobility.

Concurrent validity can be evaluated between the PDMS-2, AIMS, PEDI and PPEDI with these data. The AIMS is a qualitative evaluative instrument designed for use in infants between birth and 18 months of age. The relationship between PDMS-2 subtests and the AIMS can be examined in infants less than 18 months of age at consecutive time points. Level of agreement on age equivalent scores may help to define which test is optimal to use with infants with Pompe disease in clinical research and early intervention.

The PEDI focuses on level of independence in mobility and ADL, but research supports a high intra class correlation coefficient for concurrent validity with the GMFM, an evaluative tool designed to measure gross motor function. Low correlation coefficients were found between the PEDI and the PDMS-2 in children with mild motor impairment. Higher correlations may be found in this population with a more notable level of developmental disability. High correlation coefficients with the PEDI functional skills mobility domain would provide increased validity for use of the PDMS-2 in children with Pompe disease.

The PPEDI includes all of the items listed in the PEDI as well as 59 additional items that are designed to reflect clinically relevant skills for children with Pompe disease. Many of the additional mobility items that were added are similar to items that are on the locomotion and stationary subtest of the PDMS-2. A strong relationship in concurrent validity testing between the PDMS-2 and the
PPEDI would also provide increased validity for PDMS-2 use in children with Pompe disease.

Established instrument validity for use in Pompe disease is necessary to accurately document disease history, guide clinical trial protocol design, evaluate efficacy of the intervention, and supplement clinical decision-making.
References


