RELIABILITY AND FUNCTIONAL STATUS MEASUREMENT (ADLs) FOR CHILDREN WITH SPECIAL HEALTH CARE NEEDS (CSHCN) LIVING IN THE COMMUNITY

Charles D. Phillips, PhD, MPH
Ashweeta Patnaik MPH
James A. Dyer PhD
Emily Naiser MPH
Joshua A. Johnson MS
Constance J. Fournier PhD
Timothy R. Elliott PhD

School of Rural Public Health, Texas A&M Health Science Center
Department of Educational Psychology, Texas A&M University
Public Policy Research Institute, Texas A&M University
Word count: approximately 4,750 word in text

Running Title: FUNCTIONAL STATUS MEASUREMENT FOR CSHCN

Keywords: ADLs, inter-rater reliability, children with special health care needs (CSHCN), EPSDT, Medicaid, MDS, MDS-HC©
RELIABILITY AND FUNCTIONAL STATUS MEASUREMENT (ADLs) FOR CHILDREN WITH SPECIAL HEALTH CARE NEEDS (CSHCN) LIVING IN THE COMMUNITY

ABSTRACT

Purpose: Meeting the personal care challenges of children with special health care needs (CSHCN) in the community demands a reliable picture of their ability to perform activities of daily living (ADLs). This research investigates factors affecting the inter-rater reliability of functional assessments of CSHCN in the community based largely on caregiver reports.

Methods: The data include dual assessments for 236 community-dwelling CSHCN seeking or receiving Medicaid Personal Care Services (PCS) in a single state. They were collected in 2007 in assessments by master’s trained social workers or registered nurses using a standardized assessment instrument.

Results: The results indicate disagreements on functional function were strongly affected by the child’s underlying level of ADL impairment. Assessments of the most or least impaired CSHCN exhibited the greatest agreement. The greatest disagreement occurred when assessors faced children with moderate to moderately severe impairment. Also, specific ADLs where the greatest and least agreement occurred varied by the level of the child’s overall ADL impairment.

Conclusions: This research emphasized the importance of the measurement of CSHCN’s functional status. For a wide range of programs and services, precise measures of function are increasingly important. These results imply that the most serious measurement challenge in this arena, for researchers and service program staff,
arises when one deals with CSHCN who experience moderately or moderately severe functional impairment.
RELIABILITY AND FUNCTIONAL STATUS MEASUREMENT (ADLs) CHILDREN WITH SPECIAL HEALTH CARE NEEDS (CSHCN) LIVING IN THE COMMUNITY

BACKGROUND

Measures of a child’s ability to perform the various tasks considered the activities of daily living (ADLs) serve as major conduits of crucial information about children with special health care needs (CSHCN). Ideally, these measures provide meaningful information about the nature and extent of functional limitations [1]. The measures transmit vital information about the intensity of the disability burden the child and the child’s family face. A child’s ability to independently perform ADLs plays a major role in family’s perceptions of their need for personal care assistance [2]. ADL measures also provide critical information to caregivers, service providers, and to officials in public programs concerning a child’s need for assistance [3] that will prevent and alleviate secondary problems too often experienced by children living with a disabling condition [4].

Reliability of functional measurement is crucial for the individual child who may over the long course of a chronic illness be assessed by different individuals. As reliability decreases, the certainty that a child’s true needs will be accurately recorded and appropriately met decreases. In addition, reliable functional measurement is also important at the programmatic level. Unreliable measurement leads to both inequitable and inefficient resource allocation in service programs like those supported by the public (e.g. Medicaid or Medicare). Two children with the same needs assessed differently due to unreliable measures receive different levels of service, introducing inequity into a program. One of those two similar children receives too high a level of

WORKING DRAFT – not for attribution, citation or quotation without authors’ permission.
services, introducing inefficiency into the program. The other child is receiving too low a level of services, leaving the child and family with unmet care needs.

The route to good reliability in measurement is not shrouded in mystery. In traditional evaluations of the inter-rater reliability of measurement, one is concerned with four basic sources of variation: (#1) the stimulus itself, (#2) the specially-trained assessors’ perceptions of that stimulus, (#3) the item measuring the intensity of the stimulus with a fixed response set, and (#4) some variety of random factors. If assessor agreement (reliability) fails to achieve the necessary level, then except in the most extraordinary of circumstances, the third of these four elements will shoulder the blame.

When testing inter-rater reliability, we do not expect the stimulus to vary dramatically between assessments. If trained assessors’ perceptions of invariant stimuli vary, then the item must be at fault. Random error enters the measurement arena directly on the heels of poor item or response construction. An item may lack clarity; a response set may lack sufficient detail to assure reliability; a coding scheme may be too complex for assessors to apply consistently. The world of reliability in most test situations is quite circumscribed, and the characteristics of the items themselves are almost always at its heart [5].

When attempting to obtain valid and reliable information about the functional status of community-dwelling CSHCN, providers and public program staff face a variety of barriers not present in the world of traditional measurement. A child’s need for assistance in using the toilet room in one hospital is unlikely to differ dramatically from her or his need for assistance in using the toilet room in another hospital. But, as with adults [6], the assessment of a child’s ability to perform that function in a hospital may bear little relationship to the child’s ability to use the toilet elsewhere. The child’s ability
to use the toilet room at home, in school, or in any other community-setting in which she or he spends considerable time may vary dramatically from place to place or time to time [7]. While this example emphasizes the ability to use the toilet room, the same logic applies to all ADLs.

A wide range of factors might affect reliability estimates for functional performance measures in such settings. For example, in operational program environments in the community, assessments to test inter-rater reliability can rarely be spaced as desired. The goal in the reliability trial discussed in this research was for the second assessment to take place within seven days of the first assessment. This time span was chosen to minimize the possibility of true change in the child, changes in the household environment, or changes in those in the household providing input into the functional assessment. However, despite our best efforts, an average of 18 days elapsed between our two assessments.

Our analyses in this research focused on disagreement among our field assessors attempting to assess the functional status of CSHCN in the community for purposes of allocating Medicaid personal care resources (hours of nurse aide care). We evaluated the effect of some of these factors “external” to the traditional measurement model on the reliability of judgments concerning ADL function in this vulnerable population. Another way of considering this issue is that this research investigated a number of factors that would, in other studies, simply be assumed to introduce some nominal level of random measurement error.

Additionally, most analyses of inter-rater reliability fail to investigate the degree to which specific characteristics of the individuals being assessed may affect the reliability of the assessment, no matter what the time span between assessments might
be. However, certain characteristics of a specific subject or individual may dramatically affect the ability of assessors to agree on that individual’s status. For example, earlier research shows that the reliability of ADL assessments in the elderly in nursing homes is affected by the degree to which the resident exhibits cognitive impairment. In general, these ADL measures showed excellent inter-rater reliability [8]. The resident’s level of cognitive impairment was, however, inversely related to the reliability of their ADL assessments. Assessments of the more cognitively impaired were less reliable [9].

As noted above, greater reliability of functional status measurement can mean more appropriate and equitable distribution of program or provider resources and fewer unmet needs among CSHCN. To the degree that we understand the role of such factors as the time between assessments or the regional office in which assessors work, we gain knowledge that may allow researchers, service providers, and public program staff to formulate instrument testing procedures, assessment protocols and training regimens that assure greater reliability in functional measurement for CSHCN. To the degree that we identify client characteristics that affect reliability, providers and program managers can modify training or instrumentation in attempts to assure the allocation of appropriate care resources to these children.

**METHODS**

**Data Collection**

Data collection using the Personal Comprehensive Assessment Form (PCAF) was done as a part of regularly scheduled evaluations of the long-term care needs of children receiving or seeking Medicaid personal care services (PCS).¹ Initial assessments were

¹ Copies of the PCAF instruments, discussion of the project and previous research products generated as part of the project can be found on the internet at [blinded].
completed by fulltime case managers employed in the eleven health districts designated by the Texas Department of State Health Services (DSHS). All these case managers regularly evaluated and managed younger clients as part of their work in the Women, Infants, and Children (WIC) Program, and they were among the most experienced case managers working in their regions. As noted earlier, the second assessment was performed, on average within 18 days, by a social worker who worked on a contract basis for the DSHS as a case manager in the WIC program.

The training on the use of the PCAF instruments was 16 hours long and spread over two days. It included a detailed consideration of the instrument items, use of the instrument to assess an analogue client, and a group debriefing after this assessment. During the course of the data collection, research staff monitored the assessments and provided feedback to case managers when coding errors or other problems were discovered.

The target population for the study was children between the ages of four and 20 who were seeking or receiving personal care services under the Medicaid PCS program throughout the 11 health regions of the state. All clients being assessed by a participating DSHS case manager were invited to participate in the PCAF dual assessment process, until the target number of evaluations from a region had been obtained. The case managers evaluated those clients scheduled for their semi-annual re-assessment or clients who were just applying for PCS services. A $30 incentive in the form of a VISA gift card was paid for participation in the second assessment. Of the 287

---

2 The initial sampling plan used quotas so that the clients would be selected approximately proportionally to the Medicaid population size of the region in which they were located. The planned sample size was 264 clients aged 4-20. The planned sample size was reached, but assessments reflected current patterns of application for PCS services rather than the size of the overall Medicaid population in the region.

**WORKING DRAFT – not for attribution, citation or quotation without authors’ permission.**
contacts, only 22 declined to participate in the second assessment, giving the study a cooperation rate of 92%.

Data were collected over a three month period between November of 2007 and January of 2008. A total of 265 evaluations were obtained, however, 29 were deleted from the analysis because of high levels of missing data or the lack of a second assessment. These 265 assessments constituted roughly a 10 percent sample of all children aged 4-20 who received or sought PCS services at this time. Analyses indicate that 48% of the study households included only one parent; two parents resided in 34% of the study households; 18% of households had no parent in residence. Following each assessment’s completion, the PCAFs were sent to the Public Policy Research Institute (PPRI) at Texas A&M University. There, the forms were reviewed, and data were coded, entered, and verified.

As noted above, the PCAF for children 4-20 years of age (PCAF 4-20) was used in this effort. This assessment was purpose-built by the authors for the Texas Health and Human Services Commission, but it was also based in large measure on items, or variants of items, included in the Minimum Data Set for Nursing Home Resident Assessment (MDS) developed for the Centers for Medicare and Medicaid Services in the late 1980s [10] and the Minimum Data Set for Home Care (MDS-HC©) developed by interRAI [11]. Both the MDS and the MDS-HC© were developed for use with frail elderly and have proven reliability and validity when used with older populations [8, 10, 12-13]. Considerable evidence suggests that the MDS, when implemented nationally

---

3 interRAI is an international organization of health professionals in over 30 countries dedicated to the development of assessment instruments for vulnerable population around the world. Governmental agencies, researchers and health care providers are allowed, under agreement with interRAI, free use of all assessment and treatment planning material developed by interRAI.
improved the quality of nursing home care, with its standardized measures and focus on individualized care planning [14-19].

**Measures**

**Dependent variable.** The PCAF 4-20 used those ten items related to assistance provided in activities of daily living. The items focused on the assistance provided to participants in bed mobility, positioning while upright, eating, transfer from one surface to another, mobility inside the dwelling, mobility outside the dwelling, toilet use, dressing, personal hygiene, and bathing. All ADL items but bathing used the same response set.

The response set for each item ranged from total independence (0) to total dependence (5). The additional codes included set-up help only (1); supervision, cueing and redirection (2); limited assistance (3); hands-on assistance but only rare weight-bearing assistance; and extensive assistance (4); which denotes some involvement by the participant but regular weight-bearing help from a caregiver. The codes for bathing also ranged from total independence (0) to total dependence (5), but they differed slightly from the codes for the other ADLs. All items referenced ADL performance over a seven-day period. These items have proven both reliable and valid when used in studies of ADL function among adults [8, 10, 13, 20]. The dependent variable in our analyses was constructed by first summing the scores on each of the ten ADL measures to create an ADL score for each assessor. The difference between the ADL scores of the two assessors was then calculated, and the absolute value of this difference was used as a measure of total disagreement across all ADLs.

---

4 Positioning while upright is not an MDS item. However, it used the MDS response codes.
Independent variables. Our investigation included factors that might influence the disagreement between assessors rating a child’s functional status. The potential sources of such variance in functional status measurement are outlined in figure 1.

These include:

- differences between assessors,
- changes between the two assessments,
- characteristics of the child being assessed,
- the severity of the child’s functional impairment, and
- random error.

**Figure 1: Factors Affecting Inter-Rater Reliability of Functional Assessments of CSHCN Living in the Community**
The child’s level of cognitive function was measured on an eight point scale constructed as an additive scale for short-term memory, long-term memory, procedural task performance and cognitive skills for decision making ($\alpha = 0.75$).\(^5\)

One of our major independent variable analyses was a ten-point scale reflecting the intensity of the child’s needs for assistance with ADLs. For each ADL in which the child needed hands-on assistance, the ADL was scored as one. If hands-on assistance was not needed, then the score was zero for that ADL. The additive scale based on a child’s ratings on these ten ADL items had excellent internal consistency for both groups of assessors. ($\alpha = 0.91$ for DSHS case managers or assessors; $\alpha = 0.93$ for contract case managers).

Figure 1 also indicates how these potential sources of variance are measured in this research. Some of these variables reflect characteristics of the child being assessed that might have affected the level of disagreement among the assessors -- client age, gender, and race/ethnicity. Age was calculated in years using the assessment date and client date of birth. The presence of absence of an intellectual disability (ID) was also included in the model. If either assessor indicated that the child had an intellectual disability, then the child was coded as likely to have an ID, versus unlikely to have an ID.

Hands-on assistance was used as the cut-point for the individual ADL measures for a number of reasons. First, a score on this ADL scale is clearly interpretable to all concerned. If a child has a score of six on the scale then that child needs hands-on assistance with six ADLs. Such “transparency” is important when developing measures

---

\(^5\) A cognitive scale was constructed from each of the two assessments for a child. The scores on these scales were then averaged.

**WORKING DRAFT – not for attribution, citation or quotation without authors’ permission.**
that may be used in public programs serving a diverse clientele. Second, the bulk of PCS time is spent in hands-on care, and this approach acknowledges that reality.

Given that some assessors disagreed on a child’s ADL challenges, we separately calculated the Hands-On ADL Scale for the assessment done by the state case manager and the assessment done by the contract case manager. The best available estimate of a child’s overall ADL status was considered the average of those two scales. That average was used as the measure of a child’s overall ADL status.

The Hands-On ADL Scale was entered into the model as a categorical variable, with each category representing one quartile on the scale. The first quartile included children who received hands-on assistance in fewer than three ADLs (mildly impaired); the second quartile included children who received physical assistance in three to five ADLs (moderately impaired); those who received hands-on help in six to nine ADLs comprised the third quartile (severely impaired). Children who received physical assistance with all ten ADLs were in the fourth quartile (very severely impaired).

In our analyses, we also investigated the effects of factors related to the assessors and the assessment process. We included the number of days between the two assessments to adjust for the possibility of any real changes in the child or the child’s environment. We included the difference in the number of PCAF assessments completed by the state case manager and the number of PCAF assessments completed by the contracted case manager to adjust for potential differences in their experience with the assessment tool.⁶ We also included information on whether the state case manager, who had information from the child’s Medicaid application that was not

---

⁶ These variables were reduced to four categories for ease of interpretation. Each category represents a quartile of the sample.
available to the contract case manager, believed the child was being assessed to evaluate the effects of a purely medical problem, a behavioral health condition, or both medical and behavioral health needs.

To determine whether organizational or regional factors (Texas has regions with very distinct cultural differences) might influence the results, we used a series of dichotomous variables representing the 11 health regions in the state. We also used a series of dummy variables representing each assessor. Finally, the models included a measure indicating who might have been available in the household to provide information on the child’s condition – both parents, one parent, or some other primary caregiver.

**Analysis Strategy**

Descriptive statistics were reviewed, and then the measures discussed above were included in an analysis of variance (ANOVA). Next, the individual parameters for critical variables included in the ANOVA were evaluated, and then disagreement levels for specific ADL items were investigated. Given the number of variables included in the model and the number of parameters being estimated, only those variables or parameters where \( p < 0.01 \) were considered statistically significant and discussed. This strategy provided our results with some protection against Type I error due to multiple comparisons [20].

**RESULTS**

Table 1 presents descriptive data for this sample of children with special health care needs involved with the Medicaid PCS program in Texas. A majority of the CSHCN in our sample was members of a United States ethnic or racial minority, were male, or
had a diagnosis of some type of intellectual disability. The median age of the children was 12. Roughly half of these children presented with a medical condition alone, while the remainder of the sample was split relatively evenly between those presenting with a psychiatric, developmental or behavioral health problem alone, or with both medical and behavior health problems. Roughly half of these children could make themselves understood all of the time or most of the time. Of the ten ADL tasks investigated, half of the sample received needed hands-on assistance with over half of these activities, while almost one-quarter received hands-on assistance with all ten ADLs.

Table 1 also provides information on the settings, contexts, and environmental factors that might affect assessor agreement. In just under one-half of the homes, the primary caregiver was a single parent. On average, 18 days elapsed between the two assessments (median = 12). The assessors performing assessments on the same child usually differed in their experience with the assessment process. The average difference in the number of assessments performed was six, and the median difference was four. Thirty state case managers were involved in assessments, and eighteen contract case managers worked with the project.
### Participant Characteristics

<table>
<thead>
<tr>
<th>Variables</th>
<th>Categories</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male</td>
<td>140</td>
<td>59%</td>
</tr>
<tr>
<td>Race</td>
<td>White</td>
<td>64</td>
<td>27%</td>
</tr>
<tr>
<td>Age Group</td>
<td>4-8 years old</td>
<td>54</td>
<td>23%</td>
</tr>
<tr>
<td></td>
<td>9-12 years old</td>
<td>59</td>
<td>25%</td>
</tr>
<tr>
<td></td>
<td>13-16 years old</td>
<td>58</td>
<td>25%</td>
</tr>
<tr>
<td></td>
<td>17-20 year old</td>
<td>65</td>
<td>28%</td>
</tr>
<tr>
<td>Child's current qualifying conditions</td>
<td>Medical only</td>
<td>108</td>
<td>46%</td>
</tr>
<tr>
<td></td>
<td>Psychiatric/Developmental/Behavioral only</td>
<td>66</td>
<td>28%</td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>62</td>
<td>26%</td>
</tr>
<tr>
<td>Intellectual Disability</td>
<td>Condition diagnosed</td>
<td>134</td>
<td>57%</td>
</tr>
<tr>
<td>Cognitive Function</td>
<td>0-7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean=</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>median=</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADLs With Which Child Needs Hands-on Assistance</td>
<td>Mildly Impaired = 0 to 2</td>
<td>56</td>
<td>24%</td>
</tr>
<tr>
<td></td>
<td>Moderately Impaired =3 to 5</td>
<td>67</td>
<td>28%</td>
</tr>
<tr>
<td></td>
<td>Severely Impaired = 6 to 9</td>
<td>58</td>
<td>25%</td>
</tr>
<tr>
<td></td>
<td>Very Severely Impaired =10</td>
<td>55</td>
<td>23%</td>
</tr>
<tr>
<td>Making Self Understood</td>
<td>Understood</td>
<td>43</td>
<td>18%</td>
</tr>
<tr>
<td></td>
<td>Usually Understood</td>
<td>58</td>
<td>25%</td>
</tr>
<tr>
<td></td>
<td>Sometimes Understood</td>
<td>64</td>
<td>27%</td>
</tr>
<tr>
<td></td>
<td>Rarely/never understood</td>
<td>70</td>
<td>30%</td>
</tr>
</tbody>
</table>

### Environmental/Situational Factors

<table>
<thead>
<tr>
<th>Variables</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parents in Household</td>
<td>No Parents in Household</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>1 Parent</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td>2 Parents</td>
<td>81</td>
</tr>
<tr>
<td>Days between Assessments</td>
<td>1 to 7</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>8 to 12</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>13 to 22</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>23 to 83</td>
<td>55</td>
</tr>
<tr>
<td>Difference in Number of Assessments Conducted</td>
<td>0 to 1</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>2 to 4</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>5 to 11</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>12 to 25</td>
<td>52</td>
</tr>
<tr>
<td>Regions</td>
<td>Regions 1-11</td>
<td>N/A</td>
</tr>
<tr>
<td>Assessors</td>
<td>State Assessors 1-30</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Contract Assessors 1-18</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Table 1: Descriptive data for CSHCN with Dual Assessments (n=236)**
Table 2 presents the ANOVA results for our analysis of assessor disagreement, using our full model. As these results indicate, only one factor exhibited a statistically significant relationship with the level of disagreement: the severity of the child’s functional impairment.\(^7\) When one reviews the results of the full model that includes the assessors’ identity, the model explains slightly more than one-half of the variance in assessor disagreement on our functional status measures (\(R^2=0.54\)). Neither the other characteristics of the child (e.g., age, gender, or ethnicity) nor the characteristics of the environment or setting (e.g., days between assessments, differences in experience of the assessors) had a significant impact on the level of disagreement concerning a child’s ADL impairment.

\(^7\) Region and assessor identity were strongly collinear. Both could not be used in the same model. But, when region alone was entered into the model, it did not exhibit a significant relationship with the level of disagreement. It was differences among assessors as individuals, rather than workers in a region, that affected the level of disagreement.
When one reviews the specific parameters for the measure of ADL impairment, the results indicate that the greatest agreement among assessors was found in the assessments of those children who rated in the fourth quartile of the overall ADL scale, which included individuals designated as very severely impaired because they received hands-on assistance with all ten ADLs. The next lowest level of disagreement in the dual assessments appeared in the assessments of those children who were least impaired. It was among the assessments of CSHCN in the two middle quartiles of the ADL scale that the research team found significantly higher levels of disagreement. For those children in the second quartile (moderately impaired), the level of disagreement

Table 2: ANOVA Investigating Factors Affecting Assessor Disagreement

When one reviews the specific parameters for the measure of ADL impairment, the results indicate that the greatest agreement among assessors was found in the assessments of those children who rated in the fourth quartile of the overall ADL scale, which included individuals designated as very severely impaired because they received hands-on assistance with all ten ADLs. The next lowest level of disagreement in the dual assessments appeared in the assessments of those children who were least impaired. It was among the assessments of CSHCN in the two middle quartiles of the ADL scale that the research team found significantly higher levels of disagreement. For those children in the second quartile (moderately impaired), the level of disagreement
was approximately six points higher than in the assessments of the most impaired children. But, it was in the third quartile of the sample where the research team found the highest levels of disagreement. The assessments of those children who needed hands-on assistance with six to nine ADLs averaging over ten points of disagreement above that observed in the assessments of those most severely impaired.

| Regression Parameter                  | Estimate | Standard Error | t Value | Pr > |t| |
|--------------------------------------|----------|----------------|---------|-------|---|
| Hand-On Assistance with 2 or fewer ADLs | 3.58     | 1.84           | 1.95    | 0.0527|
| **Hand-On Assistance with 3 to 5 ADLs** | **6.067** | **1.56**       | **3.88** | **0.0002**|
| **Hand-On Assistance with 6 to 9 ADLs** | **10.18** | **1.409**      | **7.23** | **<0.0001**|
| Hand-On Assistance with all 10 ADLs  | Reference category |

Table 3: Parameter Estimates for Different Levels of Functional Impairment

Table 4 presents a more finely-grained view of the differences in assessors’ judgments. The results do not show that disagreement resulted from the scores for particular ADLs; the results do not show disagreement as basically random. Instead, the results display a very clear pattern for the levels of disagreement. For those who are only mildly impaired, agreement is lowest when assessors attempt to determine whether these children need hands-on assistance with personal hygiene and bathing. For those who are moderately severely impaired, the agreement hits its lowest points for assessments of toilet use, dressing, and locomotion outside. Bed mobility, positioning, eating, transfer, and locomotion inside generate the greatest problems in consistent evaluation for those who are severely impaired.
<table>
<thead>
<tr>
<th>Individual ADLs</th>
<th>Level of Overall ADL Impairment</th>
<th>Mildly Impaired</th>
<th>Moderately Impaired</th>
<th>Severely Impaired</th>
<th>Very Severely Impaired</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Least Agreement for Assessments of Mildly Impaired</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bathing</td>
<td></td>
<td>70</td>
<td>79</td>
<td>96</td>
<td>100</td>
</tr>
<tr>
<td>Personal Hygiene</td>
<td></td>
<td>59</td>
<td>75</td>
<td>83</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Least Agreement for Assessments of Moderately Impaired</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toilet Use</td>
<td></td>
<td>91</td>
<td>73</td>
<td>93</td>
<td>100</td>
</tr>
<tr>
<td>Dressing</td>
<td></td>
<td>86</td>
<td>68</td>
<td>96</td>
<td>100</td>
</tr>
<tr>
<td>Locomotion Outside</td>
<td></td>
<td>79</td>
<td>65</td>
<td>71</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Least Agreement for Assessments of Severely Impaired</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bed Mobility</td>
<td></td>
<td>100</td>
<td>91</td>
<td>61</td>
<td>89</td>
</tr>
<tr>
<td>Positioning</td>
<td></td>
<td>100</td>
<td>87</td>
<td>52</td>
<td>100</td>
</tr>
<tr>
<td>Locomotion Inside</td>
<td></td>
<td>96</td>
<td>90</td>
<td>57</td>
<td>95</td>
</tr>
<tr>
<td>Eating</td>
<td></td>
<td>85</td>
<td>76</td>
<td>64</td>
<td>98</td>
</tr>
<tr>
<td>Transfer</td>
<td></td>
<td>96</td>
<td>79</td>
<td>70</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 4: Severity of Impairment and Assessor Agreement Concerning Hands-On Assistance with Individual ADLs

DISCUSSION

This investigation of factors affecting the reliability field assessments of ADL function has a number of limitations. The study sample is composed of children with special health care needs seeking or receiving Medicaid Personal Care Services in Texas in 2007. These children ranged in age from four to 20 years of age; they all lived in economically distressed households; they were largely members of an ethnic or racial minority. The assessors were almost uniformly social workers without the background in functional assessment that one expects in medical professionals. These assessors also sought their information on study participants’ functional status from primary caregivers. They did not perform specific functional performance tests.

Nonetheless, the individual item data on functional status generated by this field data collection displayed good to excellent inter-rater reliability and the summary functional scales exhibited excellent internal consistency. However, as in all field studies, the two assessors disagreed in some measure about the severity of participants’
functional impairment. The task set for this research was to determine what factors affected the observed level of this disagreement.

The results were, in some sense, quite heartening. The situational and environmental factors investigated here (time between assessments, household composition, assessor experience, and assessor identity) exhibited no significant impact on the level of disagreement concerning participants’ functional status. Also, the level of disagreement was unaffected by the participants’ demographic characteristics, their ability to communicate, and their cognitive function. If supported in further research, then these results can be considered good news for researchers. Some of the factors that those engaged in field research often automatically consider random measurement error were, in fact, simply random measurement error in this study; these factors had no systematic effect on the reliability of our measures of functional status.

In addition, the results presented in table 3 were not unexpected. As that table indicated, severity of impairment and inter-rater reliability exhibited a clear curvilinear relationship. Assessors agreed most often on the functional status of those participants who were either very severely impaired or mildly impaired. On almost any functional scale, it seems, the extremes are where one can generate the greatest agreement. It is when one moves into the “muddled middle” of a functional scale or functional measure that one finds more disagreement among assessors [22].

The results presented in table 4 were both less expected and less heartening. These results indicated that the study’s raters disagreed about participants’ functional status on the basis of something other than random error. The disagreement was systematic. In this research, if the child being assessed was only mildly impaired, assessors usually agreed, but they were most likely to disagree on whether the child
needed physical assistance with personal hygiene and bathing. If a CSHCN in our sample was moderately impaired, then disagreement was most likely to be focused on the child’s need for hands-on assistance with toilet use, dressing, and locomotion outside. Bed mobility, positioning, transfer, eating, and locomotion inside were the most frequent locus of disagreement for those children who were severely impaired.

These results imply that uncertainty about a participant’s need for assistance with specific ADLs varied based both the severity of their impairment and the complexity of the specific ADL under consideration. Participants who were mildly or moderately impaired (needed assistance with no more than half of the ADLs) were clearly capable of performing less complex tasks (e.g., positioning) without physical assistance. They presented assessors with difficulty in determining their need for physical assistance in the performance of ADLs that involved multiple, ordered steps and both gross and fine motor skills (e.g., dressing, personal hygiene). Impairment that resulted in the need for physical assistance in more than half, but not all, ADLs resulted in the greatest uncertainty in determining the need for help with ADLs that were less complex. These individuals were clearly impaired in the more complex ADLs noted above, but their impairment in ADLs that did not involve ordered steps and largely depended on gross motor skills (e.g., transfer, bed mobility) lead to the greatest disagreement or uncertainty.8

The discussion of ADL function in much of the disability and rehabilitation literature treats each ADL as simply another member of the set of ADLs [23]. These results imply that approach may be somewhat simplistic when one approaches issues of

---

8 Eating “appropriately” may involve ordered steps and fine motor skills. Eating, as defined in our ADL measure, simply referred to the assistance needed for a child’s to ingest nourishment.
reliability. In other areas of inquiry such as geriatrics and gerontology, discussions of functional impairment have long recognized the idea of an “ADL hierarchy” [24].

In the study of ADL impairment in the elderly, this hierarchy often leads to distinctions between “early, mid, and late loss” ADLs [25]. Early loss ADLs are those ADLs that largely involve multiple, ordered steps and fine motor skills, like dressing and personal hygiene. Mid and late loss ADLs are those ADLs that tend to involve only gross motor skills and few, if any ordered steps. Translating the results of this research into the terminology of geriatrics, what one sees among those least impaired is uncertainty about their ability to perform early loss ADLs. For those who are more impaired, uncertainty is greatest when considering mid or late loss ADLs.

One finds this same notion emerging in studies of child development [26]. Those functional tasks mastered at the earliest age are what we have here classified as “less complex” ADLs. Eating, bed mobility, positioning, walking and transfer are those ADLs first mastered by a child. Only later will the child master, or be expected to master, what we have labeled “more complex” ADLs, which involve ordered steps (socks before shoes) or fine motor skills such as using a zipper or buttoning a coat [27].

Our results indicate that distinctions among different ADLs may be useful outside of the study of functional development or inquiries into functional decline. Distinctions among different types of ADLs may be useful in discussions of the reliability of functional measurement. However, we must await the results of further studies to determine the degree to which these results are generalizable.

Nevertheless, some of the policy implications of the current research are relatively clear. Programs like Medicaid PCS provide personal care services to individuals in the community largely on the basis of their functional status [19, 28]. As
noted earlier, to the degree that individuals with the same level of functional impairment, in this case CSHCN, receive different amounts of service, scarce programmatic resources are allocated inefficiently and ineffectively. However, the problem discussed here is not necessarily irremediable. Targeted assessment procedures and protocols could be developed and implemented either in training or in the assessment process, especially if that process is automated, to reduce uncertainty about a child’s “true” level of need.

As States or programs move to more structured allocation procedures such as case-mix allocation models, other implications arise [19, 29, 30]. Most of these more structured processes are likely to provide field workers with a “presumptive, suggested, or mandated” number of hours of service for each group of program participants. As more programs move to such models, it may become increasingly important to recognize the potential problems noted in this research. Measuring and responding appropriately to the personal care challenges of children with moderate or severe levels of overall ADL impairment may present problems that do not occur when dealing with the very severely impaired or those with only mild ADL challenges.

These results may provide those involved in studies of functionality or service delivery in the community with both some measure of assurance as well as reasons for concern. If supported by subsequent research, then researchers may find some reassurance that those factors often considered random elements in our field measurement efforts are indeed random. But, for those concerned with translating functional measurement in the community into service delivery plans, these results may indicate a need for greater sensitivity to differences among different levels or types of functional impairments when one is devising strategies for measuring and for allocating...
scarce home care resources effectively, equitably, and efficiently.
References


WORKING DRAFT – not for attribution, citation or quotation without authors’ permission.