



Evaluating Functional Assessments for Older Adults: Resource Allocation

Prepared By: David Machledt

Introduction

State Medicaid agencies and their contractors rely on functional assessment tools to make coverage and eligibility decisions for long-term services and supports (LTSS). Many states use tools developed and licensed by private entities, while others have developed and implemented their own tools. The variety of tools is staggering. One national survey identified 124 different functional assessment tools in current use across the 50 states.¹ States use assessments for eligibility determinations, service planning, and, increasingly, to inform resource allocation and quality measurement.

Elsewhere we have written about the general structure and concept of comprehensive needs assessments, the turn toward standardization, as well as some of the legal issues that can turn up during implementation of a new tool.² This paper explores the terms, techniques, and procedures states and other tool developers use to validate tools and to link assessment results to resource allocation for long-term care. We aim to identify key features advocates should understand to advocate for an allocation process is properly vetted, independent, accurate, and flexible enough to meet the needs of all Medicaid beneficiaries.

¹ MEDICAID & CHIP PAYMENT ADVISORY COMM'N, *Report to Congress on Medicaid and CHIP* 69 (June 2016), <https://www.macpac.gov/publication/june-2016-report-to-congress-on-medicaid-and-chip/>.

² See Jane Perkins, *Q&A: Using Assessment Tools to Decide Medicaid Coverage: Case Developments*, NATIONAL HEALTH LAW PROGRAM (May 2016), <http://www.healthlaw.org/publications/search-publications/QA-using-assessment-tools>.

Box 1. Common Assessment Terms

Activity of Daily Living (ADL) – One of a set of functions related to self-care that an individual may require help to complete. Commonly measured ADLs include bathing, grooming, transferring, eating and toileting.

Algorithm – A set of rules or step-by-step operations performed as a part of a calculation.

Acuity – In the long-term care context, a measure of the degree or intensity of an individual's support needs, especially related to his or her cognitive status, ability to complete ADLs and IADLs, as well as behavioral health and clinical needs.

Acuity Scale – An algorithm that measures or scores an individual based on their capacity in a particular area of functional health. Acuity scales are usually based around a set of questions or tasks and are validated against existing tests or clinical records. They would typically score an individual on a range from fully independent to fully dependent on support, and assessment tools often use a number of scales to score an individual's overall level of care needs.

Instrumental Activity of Daily Living (IADL) – One of a set of functions related to managing daily life that an individual may require help to complete. Commonly measured IADLs include cooking, shopping, managing money, managing medications, and housework.

Reliability – The degree to which an assessment tool produces consistent results across different assessors (inter-rater reliability) or different times (test-retest reliability.)

Validity – The degree to which a tool meaningfully measures what it aims to measure. There are numerous types of validity, including: ability to predict, ability to generalize across populations, ability to meaningfully distinguish between relevant groups, and ability to produce results similar to other tests or tools that measure the same thing.³

Explained Variation – The extent to which a given model accounts for the actual differences between individuals across a given population. In the assessment context, this includes variation in acuity and variation in resources required or used to satisfy one's support needs.

Multiple Roles for Assessments

Functional assessment is typically a two-step process, though in some states both steps are combined. First, an individual must meet a needs-based eligibility screen, or a "level of care," (LOC) to qualify for Medicaid LTSS. Most Medicaid home and community-based services (HCBS) programs require an institutional LOC, which means that the person's support needs would qualify them for care in a nursing facility or an

³ For detailed and digestible descriptions of different types of validity, see William M.K. Trochim, *Measurement Validity Types*, WEB CTR. FOR SOC. RESEARCH METHODS <http://www.socialresearchmethods.net/kb/measval.php> (last visited June 9, 2017.)

intermediate care facility.⁴ States typically set different thresholds and pathways for institutional LOC based on a combination of medical needs, cognitive status and behavioral/mental health, and supports required to complete basic functions of daily life, including Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs) (See Box 1). Once an individual is determined eligible for Medicaid LTSS, the second step typically involves a longer, more comprehensive needs assessment to identify what kinds and intensity of supports a person requires. The comprehensive assessment may involve a completely different tool administered by a different person or entity (such as a nurse employed by a managed care plan), or it may be contemporaneous with the LOC eligibility determination. Either way, the comprehensive assessment should solicit an individual's goals, preferences, and priorities for their support and is foundational to the development of a person-centered service plan.

Commercial proprietary assessment tools

Recent trends generally favor modular comprehensive assessment tools with standardized elements that apply across a broad range of populations and eligibility groups.⁵ Many states implement an “off-the-shelf” proprietary tool. Three of the most widely used proprietary tools are:

- The **Inventory for Client and Agency Planning (ICAP)**, an assessment in use since 1986 that measures adaptive and maladaptive behaviors and physical and cognitive function;
- **InterRAI**'s suite of tools, a collection of interrelated assessments that cover a fairly wide range of settings and populations, including home care and long-term care facilities; and
- The **Supports Intensity Scale (SIS)**, an intensive face-to-face assessment specifically designed for individuals aged 16 to 72 with intellectual and developmental disabilities. The SIS has a new version adapted for children (5+).⁶

It is not uncommon for states to make some modifications to these off-the-shelf products, so that their application varies somewhat from one state to the next.

⁴ The LOC for an Intermediate Care Facility is typically different than for a nursing facility.

⁵ A modular comprehensive assessment tool measures an individual's needs across a range of areas: functional status, cognitive function, clinical needs, goals, and so forth. Modular tools include extra questions to probe deeper into a particular area when a person responds or demonstrates a need in that area. For example, if a person answers a trigger question indicating a history of cognitive and behavioral health issues, the assessment would “trigger” extra questions to probe the nature of those problems.

⁶ AM. ASS'N ON INTELLECTUAL & DEVELOPMENTAL DISABILITIES (“AAIDD”), SUPPORTS INTENSITY SCALE INFORMATION 6 (2008), <http://aaidd.org/docs/default-source/sis-docs/latestsispresentation.pdf?sfvrsn=2>.

States roles in developing and customizing assessment tools

Some states develop their own tools to maximize flexibility and customization options. Oregon and Washington both established state-specific tools based originally on InterRAI's Home Care instrument.⁷ More recently, Minnesota has implemented an ambitious, person-centered comprehensive assessment tool called MNChoices. California is also designing its own tool for its In-Home Supports & Services (IHSS) HCBS program. Developing a state-specific model requires extensive planning and development – and money – especially if the state wants to apply its own “case mix” system to assist in budget/service allocations.

The process for developing a state-specific tool or for selecting and “customizing” a commercial tool is critical. Colorado, for example, has involved stakeholders as it considers a new assessment tool for its older adult population. The state commissioned a stakeholder task force, a website and several reports that review the pluses and minuses of a variety of state-specific and commercial assessment tools. The reports evaluate the strengths and weaknesses of well-known state-specific and proprietary assessment tools, including ICAP, InterRAI and SIS.⁸ For example, ICAP was last normed in the 1980s and may reflect outdated priorities and care expectations. SIS generally shows high scores for consistency and validity, but it is also more expensive to license, takes a long time to administer, and has been criticized for invasive questions. InterRAI's suite of tools derives from and is compatible with the Resident Assessment Instrument (RAI) used by Medicare and many Medicaid programs in nursing facilities. But the validation of InterRAI's HCBS tools is not as robust as the RAI itself, and the implementation requires an effective training and testing program for assessors. As we will see below, applying any one of these tools in a new context (such as a new state) can introduce variation that undermines the validity of the tool if it has not been properly vetted in the new context. States that select a commercial product may, to varying degrees, customize that product for their programs. As we will see below, these decisions can strongly affect the validity and effectiveness of the tool.

Anatomy of an Assessment

One can conceptualize a needs assessment as a sorting tool. It consists of a number of items (questions, observations, or short tasks) intended to gauge consistently an individual's abilities across a variety of functional topics. Responses to items covering a

⁷ Oregon changed its assessment tool for in-home services when it implemented a Community First Choice program in 2013, and this assessment was later adjusted again in 2016. *Assessment Realignment FAQs*, OREGON DEPT. HUMAN SERVS., <https://www.oregon.gov/DHS/SENIORS-DISABILITIES/DD/ExceptionsANACNA/assessment-realign-faqs.pdf> (last visited June 9, 2017).

⁸ Brittany Taylor et al., *Analysis of Instruments to Assess Support Needs of People with Intellectual and Developmental Disabilities*, HUMAN SERVS. RESEARCH INST. (Oct. 2015).

given topic, such as cognitive function or ADL ability, can be used to create scales and logic algorithms that score individuals along a spectrum for that area. Two examples are InterRAI's Cognitive Performance Scale and its ADL hierarchy, which roughly tracks the typical progression of functional loss in older adults.⁹ Ultimately, a series of these scales and logic algorithms together attempt to gauge different individuals' functional acuity and sort them on a spectrum from fully independent to fully dependent.

In some cases, the algorithm may simply be an accumulation of "points," such as in Tennessee's LOC acuity scale or the District of Columbia's Long-term Care Assessment.¹⁰ In others, such as the SIS and InterRAI tools, the algorithms sort people into different pre-defined groups with similar utilization patterns and clinical characteristics according to their responses. States may use these algorithms and groupings to determine eligibility and/or allocate resources. Specific scales can be independently validated against existing tools. For example, InterRAI's Cognitive Performance Scale (CPS) was recently updated, and researchers validated it by comparing individuals' scores on their new CPS against an already validated existing tool, the Mini-Mental State Examination.¹¹

The major proprietary tools have been tested to measure their reliability (produce consistent results) and validity (accurately test what they purport to test), though no one agrees exactly what is an acceptable standard for these terms. Some of the tools, including the SIS and ICAP, have also been "normed" so results from each individual test can be compared against a broader representative base population.¹² In a perfect world, each assessment would consistently produce results that exactly describe the supports an individual needs to achieve their goals. But in this imperfect world, algorithms and assessments have more than their share of inherent and implementation-related "noise" that produces discrepancies between the assessment results and an individual's actual need (or between needs and the actual resources used). Even tools deemed "reliable" and "valid" in the published literature only accurately describe, or "explain," a fraction of the actual variation in care needs and resource utilization across a population. Similarly, reliability tests measure consistency across a population, and considerable variation always exists for testing a single individual. In short, even the best tool, faithfully implemented, will put individuals into groups that may not meet their needs at times.

⁹ See *Cognitive Performance Scale*, INTERRAI (2014), <http://www.interrai.org/assets/files/Scales/cognitive-performance-scale-2014.pdf>.

¹⁰ *TennCare Long Term Services and Supports: A Guide to Pre-Admission Evaluation Applications*, BUREAU OF TENNCARE DIV. LTSS (2014), <https://tn.gov/assets/entities/tenncare/attachments/PAEManual.pdf>.

¹¹ John N. Morris et al., *Updating the Cognitive Performance Scale*, 29 J. GERIATRIC PSYCHIATRY & NEUROLOGY 47 (2015).

¹² *Most Frequently Asked Questions about SIS*, AAIDD, <http://aaid.org/sis/product-information/faqs> (last visited June 9, 2017). The Scholastic Aptitude Test (SAT), which sorts students by national percentiles based on their scores, is an example of a "normed" test.

For example, measures of reliability attempt to quantify the variation in scores when the same individual gets evaluated by two different assessors (inter-rater reliability) or by the same assessor multiple times (test-retest reliability). Variation stems from a number of sources, including poor judgments made by assessors, inaccurate self-reporting by the individual being assessed (including recall error, difficulty understanding questions, or motivations to under- or overstate one's capabilities), or the presence or absence of family, key caregivers, or case managers in the interview process who may have information that would alter the responses. Advocates need to understand the degree of variation inherent in any functional assessment because high variation increases the chance that an individual's score will significantly diverge from their actual needs.

Inconsistencies might also result from the structure of assessment items themselves. For example, questions typically assess the frequency and type of support received over a specific period of time.¹³ If that window, or lookback period, is three days, an individual's answers might change considerably if you asked on a Monday (when her caregiver was away over the weekend) versus a Thursday. If the assessment lookback is longer, say one or two weeks, it might smooth out the day-to-day variation in support need, but this could also increase recall errors when the respondent forgets or miscalculates the actual amount of support needed during the lookback period.

The way in which answers to questions are judged is also prone to inconsistency. For example, a person who responds to a question about toileting by saying, yes, they crawl to the toilet to get there may be marked as "totally able" even though they need assistance. Or consider this hypothetical question:

"Have you needed help to take your medications in the last three days?"
Answer: Yes/No."

In this case, individuals who only needed a single verbal reminder would give the same answer as individuals who need someone to put every pill in their mouth and help them swallow. The actual support needs for these individuals are quite different, but the assessment result would be the same. Restructuring the response options as a scale, such as "never/sometimes/frequently/always," provides for more nuanced answers, but also introduces new gray areas, such as the difference between sometimes and frequently. Also, this question may introduce uncertainty based on the assessor's or respondent's interpretation of what "support" means. Is it limited to physical support, or does it include coaching, reminders and supervision? This uncertainty in interpretation is what instruction manuals and training try to limit, but no tool completely eliminates such uncertainties.

¹³ Lookback periods vary across assessment tools. CMS's Continuity Assessment Record and Evaluation (CARE) Item Set typically uses two days. InterRAI's items use three days. The Minimum Data Set for nursing facilities uses seven days. SIS has a less structured lookback, requiring the assessor to estimate the frequency and the typical amount of daily support time needed for an item over a period of a couple months.

Finally, discrepancies between measured and actual acuity can result from factors that strongly influence actual support needs but are either not measured by the tool or not factored into the underlying scales and algorithms. For example, imagine two individuals who always need supports to bathe themselves. One has a behavioral condition that makes bathing take longer. An item measuring ADL support needs might register both individuals as totally dependent. But if the assessment never asks about the behavioral health issue, the model may not account for the extra time required to help that person.

This last type of discrepancy commonly surfaces in debates about how to weigh cognitive function. While most of the commonly used tools include items that assess the need for physical supports to complete ADLs (hygiene, locomotion, toileting, or eating) and measure cognitive function, some assessments have received criticism for neglecting or undervaluing the importance of coaching or supervision to help people continue living in their community.¹⁴ For example, an individual may need someone to provide cueing to take their pills or finish a meal, to help with shopping, or to supervise their cooking.¹⁵ Living successfully in the community often requires the ability to complete ADLs *and* IADLs, and individuals who cannot access needed IADL supports may well end up unnecessarily institutionalized.

Measuring cognitive function can also be challenging for conditions like brain injuries that commonly impact higher level cognitive abilities, such as executive decision-making and attention. Such conditions might confound assessment results because they may manifest intermittently (and, thus, outside the lookback period).¹⁶ Moreover, people with brain injuries often overestimate their own functional abilities during assessments.¹⁷ The State of New York has acknowledged these problems, twice delaying transition to a uniform assessment tool for the state's TBI waiver because the new tool scored many existing waiver recipients below an institutional LOC, rendering them ineligible.¹⁸

¹⁴ For a review of how several states address the issues of cognitive function in their functional assessments, see Janet O'Keefe et al., *Medicaid Eligibility Criteria for Long Term Care Services: Access for People with Alzheimer's Disease & Other Dementias*, ALZHEIMER'S ASS'N (May 2006), <https://www.alz.org/documents/national/Medicaideligibilityissues.pdf>.

¹⁵ For example, advocates report that Tennessee's LOC assessment does not score cueing or other forms of needed supervision.

¹⁶ A longer lookback period can also make it more difficult to accurately calculate how much assistance has been provided.

¹⁷ *Medicaid Balancing Incentive Program: Recommendations for Core Assessment Tools for Individuals with Brain Injury*, NAT'L ASS'N OF STATE HEAD INJURY ADM'RS 7 (June 2015), https://www.nashia.org/pdf/sos2015/presentations/nashia-csa-recommendations_revised_june_2015.pdf.

¹⁸ People with serious mental illnesses sometimes encounter analogous problems with LOC if their symptoms manifest more sporadically. Resolving these functional eligibility issues presents a real challenge and may require programmatic changes, such as transitioning these populations to a 1915(i) HCBS program with a lower LOC and tailoring the assessment process accordingly.

Advocates should carefully evaluate the structure of any proposed tool to consider what is and what is *not* being asked, and call for changes if necessary. In Louisiana, advocates successfully pushed the state to add new questions to the state's off-the-shelf tool after showing that the questions focused only on care received and overlooked situations where people needed supports but had no available caregiver.

Implementation of the Assessment

External methodological factors – including who conducts and participates in the assessment process – represent another source of noise in assessment results. A well-designed and adequately tested assessment instrument will only work if it is properly administered by conflict-free, trained assessors who administer the test consistently. Policies that help reduce discrepancies resulting from external factors (outside the structure of the assessment tool itself) include:

- Providing training for all assessors;
- Ensuring that assessors are equipped with adequate written instructional materials when conducting interviews;
- Lowering the total number of assessors conducting interviews;
- Requiring periodic reliability testing of assessors to promote consistent results;
- Effectively mitigating potential conflicts of interest (e.g., when assessors have ties to a provider or an MCO with a financial interest in services provided);
- Effectively explaining to beneficiaries the purpose of the assessment tool and its potential effect on their service allocation;
- Incorporating observation and task-oriented elements into an assessment to minimize problems with self-reporting bias; and
- Developing effective techniques to solicit input from family, key caregivers, or case managers in the interview (without substituting their view).

Unsurprisingly, standards vary widely across states and between tools.¹⁹ For example, ICAP should be filled out by a health professional with considerable experience working with the individual, but it need not actually include the individual in the interview. The SIS, on the other hand, requires a face-to-face interview with the individual and encourages key caregivers, family, and providers to participate in that discussion. The American Association for Intellectual and Developmental Disabilities (AAIDD), which manages the SIS, recommends centralizing training more than other assessments. AAIDD sends its staff to certify state assessors, conduct reliability tests, and “train the trainer.” It also encourages states to limit the total number of assessors in an effort to promote consistency. However, state implementation of the SIS is uneven, with inconsistencies in who actually undergoes reliability testing, what training requirements are in place, and how many assessors are used. For example, a 2013 report compared

¹⁹ C. Shirk, *Comprehensive Assessments in Home and Community-Based Services*, HILLTOP INST., 13 (2009), <http://www.hilltopinstitute.org/publications/ComprehensiveAssessmentInHomeAndCommunity-BasedServices-July2009.pdf>.

Virginia’s approach of delegating SIS assessment to over 500 case managers against Oregon’s small state-wide team of just 10 highly-trained SIS assessors.²⁰

Federal standards for assessment implementation are few and not very prescriptive. Training standards are not laid out in regulations. The 1915(i) and (k) state plan HCBS options apply the most robust protections for mitigating conflict-of-interest in needs assessment. Under these rules, assessors may not:

- Be related by blood or marriage to the beneficiary or his/her paid caregiver;
- Be financially responsible to the beneficiary;
- Be empowered to make financial or health-related decisions on behalf of the individual; or
- Have a financial interest in any entity paid to provide care for the individual, unless the state establishes they are the only willing and qualified agent in the area.²¹

Given the limited guidance from federal regulations, states can take additional steps, including establishing mandatory training regimens for assessors, requiring periodic reliability checks by independent assessors to improve consistency, and establishing robust conflict-of-interest standards, which we have discussed in detail elsewhere.²²

This last point is particularly important when states contract with managed care organizations (MCOs) to conduct the second step comprehensive needs assessment. Because MCOs are paid a flat per member per month fee, regardless of services rendered, they can have a strong financial incentive to reduce access to services. This incentive could interfere with the independence of the needs assessment. A recent report of fair hearing appeals in New York found that in the six-month period after managed care plans took over the needs assessment process from the counties, the number of decisions involving home care reductions increased six-fold.²³ Three plans accounted for the vast majority of all appeals, and the total proposed reductions would have cut the aggregate hours by some 43 percent.²⁴ In the end, fully 90 percent of the proposed reductions were prevented – partly due to court opinions in the state and

²⁰ *My Life, My Community! Re-designing Supports for Virginians with Intellectual and Developmental Disabilities Project Report*, HUMAN SERVS. RESEARCH INST. (Nov. 2013).

²¹ 42 C.F.R. § 441.730(b). States using the “sole agent” exception must provide an alternative dispute resolution process, and CMS may require states to ensure a “firewall” exists between the assessment and service provision function of the entity. U.S. DEPT. OF HEALTH & HUMAN SERVS., *Medicaid Program: State Plan Home and Community-Based Services*, 79 Fed. Reg. 2993 (Jan. 16, 2014).

²² *Id.*

²³ *Mis-Managed Care: Fair Hearing Decisions on Medicaid Home Care Reductions by Managed Long Term Care Plans, June-December 2015*, MEDICAID MATTERS N.Y. 4 (July 2016), <http://medicaidmattersny.org/cms/wp-content/uploads/2016/08/Managed-Long-Term-Care-Fair-Hearing-Monitoring-Project-2016-07-14-Final.pdf>.

²⁴ *Id.*

partly due to MCOs that either failed to show up or withdrew their request at the hearings 64 percent of the time.²⁵ This evidence suggests the importance of establishing an independent needs assessment process outside the resource allocation process (or at least an effective firewall between the MCO and the assessment unit) and of monitoring transitions to or between assessment tools in real time and on an individual basis.

Using Assessment Tools for Budget Allocations

As noted above, a well-designed and executed needs assessment tool should accurately measure each individual's support needs for community life and help identify necessary services to satisfy those needs. The needs assessment must measure individual acuity and reduce, to the greatest extent possible, any discrepancy between measured acuity and actual acuity. As noted above, this task is challenging. To complicate matters, a growing number of states are also using assessments in budgeting. Actual service costs are driven by factors that go beyond an individual's acuity, and may not register on an individual functional assessment. Nonetheless, a number of states use functional assessments almost exclusively to determine budget allocations.

Box 2. Key Definitions for Resource Allocation

Case Mix – A system that classifies individuals based on characteristics linked to the total resources actually used for their care. A case-mix grouping system, one of the most common approaches, sorts individuals into clinically meaningful groups that share similar resource use patterns.

Case Mix Index – A ratio indicating the intensity of resources used by one case mix group relative to the other case mix groups in the system.

Coefficient of Variation – Also known as relative standard deviation, this statistical measure of difference divides the standard deviation (SD) of a group by its mean. For example, a group with a mean of 100 (SD+/- 4) and a similarly sized group with a mean of 200 (SD+/- 8) would have the same coefficient of variation. In case-mix systems, it facilitates comparisons of relative internal homogeneity even when groups have different average resource utilization.

Resource Utilization Group (RUG) – One of a set of clinically meaningful categories grouping individuals with similar resource use patterns. RUGs, the most widely used grouping case mix system for long term care, are used to allocate nursing facility payments in Medicare and Medicaid. InterRAI researchers have adapted RUGs for home care settings.

While an individual's acuity logically correlates with service intensity, it is by no means the only, or even the principal, driver of systemic costs for providing those needed services. Other factors influence the actual cost of care for Medicaid, including living situation (alone, with family, or in residential facility), availability of informal supports, and geographic variation in wages, provider type and availability. So using acuity assessments to determine an individual's service budget introduces a whole new set of confounding factors to the assessment process, including broad questions about the adequacy of funding across the whole HCBS system. Resource allocation algorithms also blur the line between an *individualized* assessment of need and a population-based

²⁵ *Id.*

resource allocation/payment system.²⁶ This section briefly summarizes one common approach that links functional assessments to resource allocation: the grouping case-mix system.²⁷

Case-Mix Groups

The grouping case mix system is one of several methods for allocating LTSS resources. Some states use time estimates (for various support activities) to estimate total per diem costs. Others use weighted “point” systems to estimate an individual’s cumulative need and inform his budget. The grouping case-mix system sorts individuals into clinically meaningful categories that share similar resource use patterns. Researchers start with a diverse sample of individuals with LTSS needs, typically people already receiving services who have some sort of existing claims record. Each individual in the sample gets assessed and researchers then apply algorithms to sort the sample by functional status, clinical needs, and so forth.²⁸

Finally, researchers link the assessment data (measuring individual acuity) with data measuring the type and amount of services the individuals in the sample population actually used during the window of analysis.²⁹ Statistical analysis identifies which characteristics (such as ADL status) best predict the observed patterns of resource use. Usually the resource use is based on historical claims data for individuals in the sample, but it could derive from a time-use study or other data. Researchers then divide the sample into groups of individuals with similar resource-use patterns. The number of groups is not fixed, but rather based on how many “meaningful” divisions you can make that show both a distinct use pattern between two sets of individuals and reasonable homogeneity within each set. Not all groups are derived statistically; some may reflect clinically salient features, such as a specific diagnosis or clinical need. In some cases a single element, such as the severity of a pressure sore, can substantially impact the group an individual ends up in.

²⁶ The budget-related algorithms undergirding the actual acuity items especially blur this boundary, as they sort beneficiaries into groups and everyone in the group is treated the same.

²⁷ For an example of a detailed discussion of case-mix systems written by InterRAI researchers, see Mary L. James et al., *Design Principles for HCBS Case Mix: A Primer*, UNIV. OF MICH. INST. OF GERONTOLOGY (June 2015),

<http://www.nasuad.org/sites/nasuad/files/HCBS%20Case%20Mix%20Primer.pdf>.

²⁸ In validations of InterRAI’s case mix system, the jurisdiction was already using InterRAI’s RUG III-HC assessment tool.

²⁹ Very expensive and uncommon services, like home modifications, should be excluded from the case mix, because they skew the results. See Mary James et al., *supra* note 27, at 18.

Box 3. The Significance of Variation

To illustrate the significance of variation within case-mix groups, consider the following hypothetical, with five individuals sorted into each in four case-mix groupings.³⁰ The second column represents each individual's actual average per diem based on historical claims.

Case-Mix Group	Historical Per Diem for each individual in RUG (\$)	Group average per diem (\$)	Coefficient of Variation
Group 1	85, 90, 97, 103, 125	100	0.16
Group 2	40, 54, 66, 140, 200	100	0.68
Group 3	40, 50, 55, 60, 295	100	1.09
Group 4	170, 180, 194, 206, 250	200	0.16

In Group 1, the internal variation is low. Policymakers can be reasonably confident that setting an individual budget allocation at \$100 would roughly fit most individuals' usage based on historical average claims. But if the internal variation is higher, as in Group 2, then a \$100 average per diem will fall far short of the needs of beneficiaries at the high end of the group (\$140/day and \$200/day), while others will receive budget allocations that far exceed their historical use. In Group 3, a single unexcluded outlier with extremely high costs would face a dramatic cut in services (from \$295 to \$100/day), while everyone else in the group would be authorized per diem budgets substantially exceeding their need, based on historical use. Group 3 illustrates why it is important to identify and separate outliers. Researchers can review what made an outlier so much more expensive and either create a separate RUG for that type of individual or exclude that individual or service from the case mix system and pay for it separately.

Note that Group 4 has the same variation as Group 1, but has a twice the absolute range in resources used (\$80 vs. \$40). This is a feature of the statistical measure of coefficient of variation. It suggests that variation between individuals in groups with relatively high needs may be greater in absolute terms (and still be considered "acceptable").

After sorting individuals into groups, researchers calculate the average amount of resources used by each group based on available utilization data. Each group's average then gets compared against the average resource utilization in other groups, creating a relative ratio for each group. This ratio is known the case-mix index (CMI). Because CMIs are ratios, they are not tied to any particular absolute resource use level. They shape how available resources get divided up across a population, but certainly

³⁰ Note that InterRAI's published validations have RUGS with anywhere from 2 to 8000 individuals in each RUG. Larger samples decrease the effect of individual outliers.

do not ensure the HCBS system itself is sufficiently funded or that each individual is getting enough funding to cover what she needs.³¹

To be effective, a case-mix grouping system also relies on an assessment tool sensitive enough to flag outliers. Individuals (or services) likely to have unusually high costs should be considered outliers and should either be separated into their own groups or excluded entirely from the case-mix system (See Box 3). If not excluded, outliers can throw off the proportions of resources dedicated to any single case-mix group. Unfortunately many tools are not very capable at identifying individual outliers. One study of four state assessment systems found that overall, the tools only identified between 24% to 37% of the highest cost users.³⁵ While an individual's needs may vary from year to year, especially based on medical conditions, it is not hard to see how a case-mix system that fails to identify a majority of high needs individuals could miscategorize certain individuals who can be expected to have unusually high needs with the result that they get allocated resources inadequate to meet their needs.

Box 4. The RUG Long Term Care Case-Mix System

Perhaps the most widespread example of a grouping case-mix is the Resource Utilization Group (RUG) system, which has been used to since 1998 as the basis of Medicare's prospective payment system for skilled nursing facilities and is widely used in Medicaid.³² It is now on its fourth iteration (RUG IV).³³ In the 1990s a group of researchers (now InterRAI) adapted the RUG system – and its associated assessment tool, the Resident Assessment Instrument – to apply in home care settings. The resulting RUG III – Home Care (RUG-III/HC) has become a widely used case-mix grouping system for older adults receiving Medicaid home and community based services (HCBS). The RUG IV system can have as many as 66 categories, but the RUG-III/HC algorithm consists of just 23.³⁴

³¹ Case-mix system designers acknowledge this shortcoming of this resource allocation tool. Mary James et al., *supra* note 27, at 6.

³² Over half the state Medicaid agencies use RUG III for nursing facility payments, though they can adapt slightly different versions according to their needs. *Long-Term Care Facility Resident Assessment Instrument Version 3.0 User's Manual: Version 1.14*, U.S. CTRS. FOR MEDICARE & MEDICAID SERVS. ("CMS") 6-2 (Oct. 2016), <https://downloads.cms.gov/files/MDS-30-RAI-Manual-V114-October-2016.pdf>.

³³ *Id.*

³⁴ A version of the RUG III/HC algorithm is available on InterRAI's website. See, *RUG-III/HC Home Care Classification*, INTERRAI, http://www.interrai.org/assets/files/Case-Mix%20Classification/rug-iiihc_23_diagram.pdf (Last visited June 9, 2017).

³⁵ Brant E. Fries et al., *Increasing Access to Home and Community Based Services through Equitable Resource Allocation Methodologies: Final Project Report*, UNIV. MICH. 18 (June 2015).

Explaining Variation

To measure the validity of a grouping case mix, researchers gauge the relative costs for individuals both *within* and *between* the different groups. The key measure for validity is the “explained variation,” or the degree to which the grouping system can account for actual differences in resources used by the sample population. A higher explained variation suggests a better fitting model, but no model will achieve, or even approach, 100% explained variance. The well-tested RUG-III case-mix for nursing facilities explained about 55% of the variation in actual nursing facility resident costs (see Box 4).³⁶ Earlier validations of InterRal’s RUG-III/HC algorithm in Michigan and Ontario showed that model accounted for 34% to 37% of the actual sample variation.³⁷ Experts consider these validations acceptable; the RUG-III/HC adequately sorts people into similar groups with similar resource patterns. But the home care model is clearly not as robust as the nursing facility RUGs and leaves considerable variation unexplained. In layman’s terms, a poorer fitting model makes it more likely that results from an individual’s acuity assessment will diverge from that individual’s actual needs or from her historic resource utilization patterns.

The degree of variation specifically *within* each group is typically described by the coefficient of variation. If the classification system is good, resource patterns within a group should be relatively homogenous, e.g., people in the group will have very similar needs. Low variation is especially important if the case-mix grouping will be used later to inform individual resource allocations (See Box 3). Historically, case-mix systems have been used to adjust payments for nursing facilities so facilities with higher needs residents receive more funds. In that instance, variation within or between particular groups can be smoothed out at the facility level. One resident may have higher needs than the assessment found, but other residents may have fewer needs than the assessment indicated. The facility could thus adjust for its residents’ individual needs and the payments might balance out. In the HCBS world, using a case-mix system for payments to a home health agency or a community-based residential provider serving a number of Medicaid beneficiaries may similarly allow the provider to adjust for individual variation among its clientele.

However, using the relatively imprecise case-mix system to determine an individual’s budget is more problematic. Remember, the allocation for each group is based on the average utilization across that group. Every group has a degree of built-in internal

³⁶ The RUG classification was developed from extensive staff time management studies. For example, the most recent RUG-IV version is based on a study from 205 nursing homes across 15 states. The study included clinical information from nearly 10,000 residents linking their health status, medical conditions, services received, and the staff time and facility resources used to meet their care needs. *Time Study (STRIVE)*, CMS, <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/SNFPPS/TimeStudy.html>, (last visited June 9, 2017).

³⁷ Magnus A. Björkgren, *A RUG-III Case-Mix System for Home Care*, 19 suppl. 2 CANADIAN J. ON AGING 106 (2000); Jeffrey W. Poss, et al., *Validation of Resource Utilization Groups Version II for Home Care (RUG-II/HC): Evidence from a Canadian Home Care Jurisdiction*, 46 MED. CARE 380 (2008).

variation, suggesting that even in the best case scenario, individuals whose historical needs were above average for the group will be short-changed, while others will be authorized more services than they may actually need. Such a one-size-fits-all budgeting process is not individualized and thus may not incorporate the principles of person-centered planning. This is why advocates should push to ensure that any individual budgeting includes mechanisms to reassess and, if necessary, adjust individual budgets. This could include having the assessment point to a range of dollars or hours rather than a single figure. Implementing a clear exceptions process is also important, especially to identify outliers that the assessment process missed.

In addition, as noted above, many factors beyond an individual's acuity strongly affect resource utilization patterns. For example, if an individual desires to self-direct their services, the hourly cost of those services is often lower. Whether and how to account for these lower costs raise important questions about equity and individual choice.

Unexplained Variation

As explained above, the characteristics and implementation of the assessment structure can create a discrepancy between a beneficiary's measured acuity and her actual acuity. In a case-mix model, that discrepancy gets compounded by additional non-acuity related factors that shape the costs of providing needed care. For example, in one validation of the SIS case-mix system, 28.8% of the variation in resource use could be explained by living arrangement (whether individuals lived alone, with family, or in certain residential settings), while the measured acuity level from the needs assessment explained just 18.8%.³⁸ In other words, more of the costs of delivering care were related to where the person lives than what their actual support needs were. For example, a person living in a small residential group home may cost significantly more to care for than another person with the same support needs who lives at home and self-directs their care. Other non-acuity related factors influencing resource use include:

- The amount of informal care received;
- Geographic variation in wages;
- Variation in provider availability;
- Prevalence of self-directed care;
- Variation in relative costs of different providers or home care agencies; and
- Changes to the service plan that do not result in a reassessment.

³⁸ Madeleine Kimmich et al., *Developing Individual Budgets and Reimbursement Levels Using the Supports Intensity Scale*, HUMAN SERVS. RESEARCH INST. App. B-7 (Apr. 2009), <http://www.hsri.org/publication/developing-individual-budgets-and-reimbursement-levels-using-the-suppo/>; Mary L. James et al., *supra* note 27, at 7.

Varying case-mix allocations based on non-acuity related factors like living situation or availability of informal supports raises important questions about individual choice. For example, if a state or MCO gives less funding to a beneficiary living with family than it does to a beneficiary with similar needs living alone, does it effectively compel the family to provide unpaid supports to make up the gap? Federal regulations clearly prohibit states from compelling unpaid informal supports.⁴² Litigation in Washington has struck down regulations that automatically reduced the allocated hours for children in the state's waiver for children with

intellectual or developmental disabilities who lived with their parents, because the state automatically assumed that the parents would meet the child's needs.⁴³ The SIS budget allocation system varies resource allocation based on living arrangement – putting adults living in residential settings in a separate category from adults with similar needs living at home. However, this method of categorizing raises the question of the extent to

Box 5. Incorporating Informal Care

Most of the LTSS in the U.S. is provided by unpaid caregivers.³⁹ The accuracy of a case-mix system depends largely on how case-mix developers measure and weigh unpaid, informal supports. Accurately measuring the amount of informal care provided, and assigning a “correct” value, is notoriously difficult and is another source of noise. Both of the most commonly cited validations of the RUG-III-HC system incorporate the relative value of unpaid, informal care provided to individuals in the sample. This factor increased the explained variation in both studies.⁴⁰ In other cases, though, including informal care did not improve the model.⁴¹

But even if a case mix model accounts for informal care, the informal care raises questions in budget allocation. Essentially, the “average” allocation in a case mix group assumes an average amount of unpaid informal care. If a state allocates all individuals in a case-mix group the same amount, individuals who have above-average access to informal care would get allocated relatively more resources, while individuals with little or no access to informal supports would receive inadequate resources to meet their needs.

³⁹ Erica L. Reaves and MaryBeth Musumeci, *Medicaid and Long-Term Services and Supports: A Primer*, HENRY J. KAISER FAM. FOUND. 1 (Dec. 2015),

<http://www.kff.org/medicaid/report/medicaid-and-long-term-services-and-supports-a-primer/>.

⁴⁰ Jeffrey W. Poss et al., *supra* note 37, at 383 (Apr. 2008) (including informal care raised explained variance from 20.5% to 37.3%); Magus A. Bjorkgren et al., *supra* note 37, at 120 (including informal care raised explained variance from 26.3% to 33.7%).

⁴¹ Mary James et al., *supra* note 27, at 35.

⁴² Regulations for 1915(c), (i), and (k) HCBS programs clearly define natural supports as “unpaid supports that are provided voluntarily to the individual in lieu of state plan HCBS.” 42 C.F.R. §§ 441.301(c)(2)(v), .540(b)(5), .725(b)(5). Decisions about informal caregivers’ role in providing services for an individual should be included in the care planning process.

⁴³ *Jenkins v. Dep’t. of Soc. & Health Servs.*, 160 Wash.2d 287, 308, 157 P.3d 388 (2007); *Samantha A. v. Dep’t. of Soc. & Health Servs.*, 171 Wash.2d 623, 256 P.3d 1138 (2011); see also *Samantha A. FAQs*, DISABILITY RIGHTS – WASH., <http://disabilityrightswa.org/advocacy-news/samantha-faqs> (last visited June 12, 2017.); see generally WASH. ADMIN. CODE § 388-106-0130 (detailing current system for allocating informal supports in CARE tool).

which living at home has historically been less expensive due to unpaid informal supports substituting for Medicaid-funded home care and many times being compelled to do so? Is there a mechanism in the budget allocation system to adjust the resource level so an adult with no available informal supports can still choose to live at home?

Finally, is the individual choosing where and how they live *first*, and then receiving the resources needed to support those decisions, or is the service allocation process instead shaping (or coercing) their decisions about where and how to live? The rules defining a home and community-based setting require that individuals may choose between available settings options, including at least a non-disability specific option. But if the available options have different costs, is the individual still offered free choice between these settings? If they currently live in a less expensive setting, such as a large group home, are there features of the process that might lock them into that setting and create barriers to choosing something else, even if it were more expensive?

Importing and Customizing Assessment and Case-Mix Tools

Developing a state-specific assessment tool and case-mix system is a challenging and expensive proposition. To soften the budget impact, states frequently borrow from their peers or purchase off-the-shelf proprietary tools. As noted above, this could include adopting a commercial assessment tool, some of which include a case-mix grouping system. With an existing assessment tool in hand, the state has to make choices about how much it will “customize” the tool and case-mix to its own circumstances. For example, the state could conduct a full analysis to create its own case mix indices or choose to import the published case-mix ratios from another system. While importing and adapting a tool may be faster and cheaper, it can also increase noise. The more “off-the-shelf” components a state uses, the more likely it is to compromise precision. Put another way, the short-cut may result in substantially lower explained variation, indicating that the case-mix ratios are not as close a “fit” for the new state.

Remember that case-mix ratios derive from specific sample populations linked with resource utilization data over a given window of time. One of the major validations for RUG-III-HC used 2003-2004 data from an LTSS system in Ontario.⁴⁴ At least one state applied the Ontario case-mix indices to roll out its own case-mix system.⁴⁵ However, if the services included in the Ontario validation case do not match closely with the services in State X – for example if one included skilled nursing and durable medical equipment, but the other only included home care – the case mix ratios might lose accuracy in State X. Similarly, if the sample used in Ontario was limited to older adults, while State X wanted to apply the case-mix system to all adults with a disability who need home care, the ratios could get thrown off. Another factor to consider is time. HCBS services have evolved considerably over the decades, and the Ontario validation is based on data nearly 15 years old. Have the services changed in the interim? Are newer social services now available that could affect resource use patterns? Finally, as mentioned above, if the new state has chronically underfunded its HCBS system,

⁴⁴ Jeffrey Poss, *supra* note 37.

⁴⁵ Arkansas imported the Ontario CMI when it rolled out its RUG III/HC model in 2015.

applying a case-mix index from another state will *not* cure that problem – it will simply reallocate insufficient funds differently.

If a state chooses to implement an assessment tool along with a case-mix system, it should involve stakeholders in a transparent decision-making process and should document how it establishes the validity of its particular version based on the decisions it made and the quality of data that supports its tool. Too many compromises without adequate testing call into question the validity and the fairness of the whole tool, even if it is a widely used commercial tool. And it should be abundantly clear that significant variation at the individual level is inherent to all these assessments.

Such complexities underscore why assessment tools should not be the sole source of an individual's budget determination, as needs assessment alone accounts for only a fraction of the variation within or between groups. Advocates should push their states to build in flexibilities that acknowledge the inherent shortcomings of their assessment and allocation models. For example, states and MCOs administering assessments and case-mix systems should recognize the implicit uncertainty in applying a population-derived budget category to a particular individual and build in mechanisms to adjust the service budget for each beneficiary as necessary.

Additional mechanisms that can improve flexibility in individual allocations include:

- Asking the state to set the budget allocation as a limited range (e.g., +/- 10 percent), allowing the person-centered planning process to accommodate individual differences in needed services or in the availability of informal care;
- Creating an exceptions process with due process rights for individuals who believe they require more services or have been improperly assessed; and
- Grandfathering existing service authorization levels for individuals so that service hours cannot be reduced absent a measureable change in functional status.

Finally, advocates should get answers to three key questions for any state in the process of adapting a tool or developing its own assessment and case-mix system:

- **How is the state incorporating informal care into its model?** States need to ensure that allocations are not directly or indirectly pressuring informal caregivers to cover hours the system will not pay for and guarantee that allocations are not systematically disadvantaging individuals who have no access to informal supports to help them stay in the community.
- **If the state is using its own historical data to develop or adapt a case-mix tool, how is it ensuring that preexisting systemic biases in HCBS funding get remedied and the overall system is adequately funded.** Remember, a case-mix system derives from historical utilization, not necessarily what utilization

*ought to be.*⁴⁶ If your state has long underfunded behavioral health services, for example, and it uses its claims data to create its own case mix ratios, that legacy of underfunding will be reflected in the new ratios.⁴⁷

- **What is the process to identify and exclude individual outliers, such as people who require expensive home modifications, from the case mix system?** If a state chooses to use a case-mix model for budget allocation, establishing a robust exceptions process is essential to avoid distortions and to ensure that the tool does not systematically disadvantage high-needs individuals.

Conclusion: Managing Transitions

Advocates should note that a crucial component of transparency in the results of needs assessments includes the underlying algorithms a state uses to weigh beneficiary responses. States should be able to provide the evidence the state collected to show that any commercial tool has been validated locally and should identify any steps the state took to adapt the tool to unique circumstances in that state. Evidence includes reliability tests of state assessors, pilot tests comparing the performance of a new tool against the former tool to illustrate who may be harmed during the transition, and any analysis the state made to apply a case-mix, particularly if the case-mix index is imported from a published study. This extra information can be difficult to obtain, but it may be critical to understanding how the state uses assessment results to inform an individual's allocated budget.

If your state is proposing to implement a new needs assessment tool, urge officials to:

- Create a robust and transparent public process to seek early input and ongoing feedback from beneficiaries and advocates;
- Establish a “grandfathering” clause to protect services and eligibility for existing beneficiaries unless the state has established clear evidence of a change in the individual's health or functional status;
- Establish a grace period within which, during the implementation phase-in, an individual can easily dispute a change due to reassessment and retain her current services. This creates a cushion for the state to work out kinks without harming individuals;
- Establish robust assessor training standards with regular reliability testing to promote consistent results between assessors;

⁴⁶ Mary L. James et al., *supra* note 27, at 6.

⁴⁷ *Id.*

- Promote the concept of independent assessments through strong conflict-of-interest standards – including standards that apply to managed care organizations – that mitigate potential financial incentives to game the assessment process;
- Seek ongoing feedback on the assessment and allocation process through an advisory committee with meaningful representation from key stakeholders including beneficiaries and their advocates;
- Demand that your state make needs assessment and case-mix results and algorithms publicly available;
- Require that beneficiaries have access to clear and understandable information about the assessment process, including who can or should be involved, what information assessors may inquire about or access, and how to interpret results;
- Ensure that notices related to the assessment process include all necessary information to meet the requirements of due process. More details are available in previous NHeLP briefs.⁴⁸

⁴⁸ Jane Perkins, Q&A: *Using Assessment Tools to Decide Medicaid Coverage: Case Developments*, NAT. HEALTH LAW PROG. (May 2016), <http://www.healthlaw.org/publications/search-publications/assessment-tools-enrollees>; Jane Perkins, *Ensuring that Assessment Tools are Available to Enrollees*, NAT. HEALTH LAW PROG. (Sept. 2015), <http://www.healthlaw.org/publications/search-publications/assessment-tools-enrollees>.