MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 3622

Corresponding Measures:

De.2. **Measure Title:** National Core Indicators for Intellectual and Developmental Disabilities (ID/DD) Home- and Community-Based Services (HCBS) Measures

Co.1.1. **Measure Steward:** Human Services Research Institute

De.3. **Brief Description of Measure:** National Core Indicators for Intellectual and Developmental Disabilities Home- and Community-Based Services Measures ("NCI for ID/DD HCBS Measures" hereafter) originate from NCI(R) In-Person Survey (IPS), an annual multi-state cross-sectional survey of adult recipients of state developmental disabilities systems´ supports and services. First developed in 1997 by the National Association of State Directors of Developmental Disabilities Services (NASDDDS) in collaboration with Human Services Research Institute (HSRI), the main aims of NCI for ID/DD HCBS Measures were to evaluate person-reported outcomes and assess state developmental disabilities service systems performance in various domains and sub-domains accordingly. The unit of analysis is "the state", and the accountable entity is the state-level entity responsible for providing and managing developmental disabilities services. Currently, 46 states and the District of Columbia are members of the NCI program. To align with member states´ fiscal schedules, the annual survey cycle typically starts on July 1 and ends on Jun 30 of the following year.

Gathering subjective information and data from people with ID/DD poses unique challenges due to potential intellectual and developmental limitations experienced by the population. As such, extensive work went into the processes of developing NCI IPS administration methods, survey methodology and measure design and revisions. The original development built on direct consultation with members of the target population and their advocates, as well as extensive literature review and testing.

The NCI for ID/DD HCBS Measures consist of 14 measures in total, including:

Five measures in the HCBS Domain: Person-Centered Planning (PCP) and Coordination

#PCP-1 The proportion of people who express they want a job who have a related goal in their service plan (Community Job Goal)

#PCP-2 The proportion of people who report their service plan includes things that are important to them (Person-Centered Goals)

#PCP-3 The proportion of people who express they want to increase independence in functional skills (ADLs) who have a related goal in their service plan (ADL Goal)

#PCP-4 The proportion of people who report they are supported to learn new things (Lifelong Learning)

#PCP-5 The proportion of people who report satisfaction with the level of participation in community inclusion activities (Satisfaction with Community Inclusion Scale)

Four measures in the HCBS Domain: Community Inclusion

#CI-1 The proportion of people who reported that they do not feel lonely often (Social Connectedness)

#CI-2 The proportion of people who reported that they have friends who are not staff or family members (Has Friends)

#CI-3 The proportion of people who report adequate transportation (Transportation Availability Scale)

#CI-4 The proportion of people who engage in activities outside the home (Community Inclusion Scale)

Four measures in the HCBS Domain: Choice and Control

#CC-1 The proportion of people who reported they chose or were aware they could request to change their staff (Chose Staff)

#CC-2 The proportion of people who reported they could change their case manager/service coordinator (Can Change Case Manager)

#CC-3 The proportion of people who live with others who report they can stay home if they choose when others in their house/home go somewhere (Can Stay Home When Others Leave)

#CC-4 The proportion of people who report making choices (independently or with help) in life decisions (Life Decisions Scale)

And one measure in the HCBS Domain: Human and Legal Rights

#HLR-1 The proportion of people who report that their personal space is respected in the home (Respect for Personal Space Scale)

1b.1. Developer Rationale: In the past 30 years, we have witnessed significant changes in the nature and extent of the publicly financed systems of services and supports for individuals with developmental disabilities and their families. Growing service needs, declining state revenues, mounting pressures on federal and state Medicaid budgets, and heightened federal quality management expectations are placing increasing demands on state developmental disabilities agencies to improve accountability, more effectively track outcomes, and strengthen their capacity to assess and improve service quality and responsiveness. The initiation of the National Core Indicators collaboration 20 years ago was in direct response to these and other forces described below. Public systems that provide home and community-based services for people with intellectual and developmental disabilities are highly diverse and widely dispersed. As the places where people live and work and the supports that they receive become more individualized, the necessity for effective performance appraisal and quality improvement systems which directly capture people’s experience of services and outcomes has continued to grow. The administration of the Home and Community Based Medicaid Waiver program by the U.S. Department of Health and Human Services Centers for Medicare and Medicaid Services (CMS) requires to states to monitor and provide assurances of the quality and performance of their support systems. Federal expectations included in the revised formal Waiver application require that each state agency explain the operational features of its planned HCBS Medicaid Waiver program in detail. States also describe the components of the comprehensive quality management strategy that they intend to employ to monitor and improve the accessibility and quality of services offered to Waiver participants and to demonstrate that required Waiver assurances are being met. Through this approach, CMS has information allowing assessment of a state´s quality management strategy based on the presence or absence of objective, measurable data demonstrating that the state is actively managing the quality of Medicaid Waiver services and supports through a continuous process of discovery, remediation, and improvement. To respond to the emphasis on performance and outcome data, CMS has recommended that states consider the use of standardized tools, such as the National Core Indicators, to gather and analyze information on Waiver participants´ outcomes and satisfaction with the services they receive. The changing expectations on the part of CMS regarding quality monitoring and performance measurement parallel gubernatorial and legislative initiatives in several states aimed at increasing government accountability, improving service outcomes, and strengthening program responsiveness.

As PRO-PMs, these measures are designed to assist states in moving beyond compliance to assess service quality, in line with the minimum standards of the federal regulations. The full measure set is designed for states to measure their performance in the HCBS Domains address by the NQF 2016 report and establish goals and benchmarks for performance and leading states to focus on quality improvement, a step beyond compliance.

The 14 measures being submitted collectively addresses 4 of the measurement domains identified by the national, multi-stakeholder Committee on HCBS convened by NQF as areas where “performance measures are needed to drive systems change, tie performance to outcomes, allow consumers to make informed choices, and compare the effectiveness of different models of HCBS and of HCBS versus institutional services.” (NQF, 2016, p. 2).

HCBS Domain: Person-Centered Planning (PCP) and Coordination (Measures #PCP-1 – #PCP-5)

CMS final rules on HCBS Waiver Requirements, published in 2014, promulgated expectations that waiver states establish a Person-Centered Planning Process and the Person-Centered Plan within a transition period now extended until 2023. The 5 measures in the PCP Domain submitted for NQF endorsement have been developed in response to state DD agencies’ requests for measures which can demonstrate the required compliance with both the person-centered planning process, and the HCBS Settings requirements. During annual meetings of the NASDDDS members (State Directors of Developmental Disabilities Services), requests were made to develop and /or adapt existing measures to assist states in determining their performance related to these new rules.

HCBS Domain: Community Inclusion (Measures #CI-1 – #CI-4)

The Americans with Disabilities Act (ADA) of 1990, and subsequent legal actions such as the 1999 Supreme Court decision in the landmark Olmstead case, have established full integration with society as a basic human right for individuals with disabilities. Thus, the adequacy and quality of state DD services depends on the extent to which they support their consumers’ community inclusion. The 4 measures in the Community Inclusion Domain submitted for NQF endorsement measure the success with which state service systems support individuals with IDD in participating in the activities outside their home and integrating with their communities. Given that community inclusion was identified by the NQF Committee on HCBS as one of the key outcomes expected of a service system, these measures are also vital in comparing competing HCBS models.

HCBS Domain: Choice and Control (Measures #CC-1 – #CC-4)

The main purpose of HCBS is to promote and maintain the wellbeing and quality of life of people with functional limitations. Ability to make life choices such as where and with whom to live, services they receive, and how those services are delivered, have been shown to improve the quality of life among people with disabilities (Willis, Grace, & Roy, 2008; Stancliffe et al., 2011). Therefore, measures that assess the level of choice and control that individuals with IDD have over their life and their services are crucial performance measures for state service systems. Monitoring changes in their measures of choice and control will help state service systems achieve continuous quality improvement. The measures are also vital in comparing and improving HCBS systems.

Stancliffe, R.J., Lakin, K.C., Larson, S., Engler, J., Taub, S., & Fortune, J. (2011). Choice of living arrangements. Journal of Intellectual Disability Research, 55(8), 746-762.

Willis, D., Grace, P. J., & Roy, C. (2008). A central unifying focus for the discipline: Facilitating humanization, meaning, choice, quality of life, and healing in living and dying. Advances in Nursing Science, 31(1), 28-40.

HCBS Domain: Human and Legal Rights (Measure #HLR-1)

Within this domain, we are submitting a measure of the respect for one’s privacy. This is an important component of basic human rights and dignity that all members of society are entitled to. The Convention on the Rights of Persons with Disabilities promulgates the protection of basic human rights for people with disabilities. Therefore, the extent to which individuals with IDD report that their personal space is respected is an indicator of the ability of an HCBS system to protect its consumers’ basic human rights and dignity. It is, therefore, a crucial measure of service quality. Monitoring changes in their scores on this measure will enable state service systems to make the necessary changes to their services to avoid intrusiveness and instead facilitate privacy and dignity among individuals with disabilities.

S.4. Numerator Statement: The NCI for ID/DD HCBS Measures use values between 0 and 1 as the scores. Typically, the numerator is the number of respondents who selected the most positive response category (e.g., "yes", "always"). The attached file SuppTable_Measures_210420_508.xlsx lists what constituted the most positive response categories for each measure item, as well as other detailed information as relevant for S.2b.

S.6. Denominator Statement: For each measure, the denominator is the number of respondents (adult recipients of state developmental disabilities services) who provided valid answers to the respective survey question, except those that meet the exclusion criteria (see S.8. below for details).

If the denominator for a state is fewer than 20, the measure score is censored to protect the confidentiality of respondents.

S.8. Denominator Exclusions: At the end of Section I, the surveyor assesses whether the respondent appears to understand at least one question and answers in a cohesive manner. This assessment is the only subjective process in the exclusion determination process, but it is not done on an arbitrary or state-by-state basis. Rather, it is based on a protocol, included in the survey manual and reviewed during surveyor trainings, that apply uniformly to all surveyors across different participating states. The protocol is straightforward—the section must be marked “valid” if at least one question in the section was answered in a manner that the basic level of comprehension was shown, and a clear response given either verbally (e.g., yes/no) or non-verbally (nodding/shaking head). NCI and participating states routinely conduct surveyor training and surveyor shadowing and reviewing processes that ensure, among other things, that surveyors are applying this assessment (whether or not Section I was valid) strictly based on the protocol. If the surveyor´s assessment is that Section I is not valid, the respondent´s Section I data are flagged for exclusion from the numerators and denominators. However, the individual is not removed from the dataset.

If Section I data are excluded, Section II data are flagged for exclusion from the numerators and denominators -unless- a proxy respondent was used in Section II. If the respondent or proxy did not answer any questions in Section II, the survey is removed from the denominators of Section II items.

Responses are excluded from numerators and denominators for Section I items if:

(a) The surveyor indicated that the respondent did not give consistent and valid responses, or

(b) All questions in Section I were left blank or marked "not applicable" or "don’t know".

Responses are excluded from numerators and denominators for Section II items if:

(a) the individual receiving supports was marked as the sole respondent to all questions in Section II but Section I was deemed invalid, or

(b) All questions in Section II were left blank or marked "not applicable" or "don’t know".

For each measure item, missing responses and responses indicating "not applicable" or "don’t know" were excluded from denominators. The distribution of exclusions among states is shown in Testing Attachment 2b2.2. Please see S.9. for more details on denominator exclusions.

De.1. Measure Type: Outcome: PRO-PM

S.17. Data Source: Instrument-Based Data

S.20. Level of Analysis: Population: Regional and State

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?

Preliminary Analysis: New Measure

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

Evidence Summary

- This is a new outcome PRO-PM measure at the population (regional and state) level that aims to assess the performance of Intellectual and Developmental Disabilities (ID/DD) Home- and Community-Based Services (HCBS) in various domains and subdomains based on National Core Indicators (NCI).
- The developer provides a logic model to depict how HBCS lead to improvements in service quality, how the 14 measures fit into the quality measurement domains, and how the measurement domains result in desired outcomes of HBCS.
- The developer demonstrated the value and meaningfulness of the measure by noting that the measures reflect priorities identified by individuals with intellectual and developmental disabilities, which were identified by experts who were convened in 2016 to establish the NQF framework for HCBS quality. Researchers at University of Minnesota's Research and Training Center on Outcome Measures conducted a study to assess the validity of the NQF Framework and determined that the content of the NQF HCBS Quality Measurement Framework had a high level of support from the stakeholder group that evaluated the content, which included individuals with ID/DD.
- The developer provided evidence demonstrating how the reporting of NCI submitted measures across various states and regions led to improved outcomes for HCBS recipients.

Question for the Committee:

○ *Is there at least one thing that the provider can do to achieve a change in the measure results?*

○ *If derived from patient report, does the target population value the measured outcome and finds it meaningful?*

Guidance from the Evidence Algorithm

Box 1: Measure assesses performance on a PRO → Box 2: Relationship between measured PRO and service is identified and supported by stated rationale → PASS

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer reports performance data for the 14 individual measures within the “Table 2a2.3 and 1b.2” of the “SuppTable”.
    - Community Job Goal:
        - 34 States with 3,282 patients
        - Mean=0.3806 with interquartile range (IQR) of 0.181
    - Social Connectedness:
        - 37 States with 11,933 patients
        - Mean=0.8826 with IQR of 0.044
    - Has Friends:
        - 37 States with 12,331 patients
        - Mean=0.7868 with IQR of 0.085
    - Chose Staff:
        - 37 States with 16,509 patients
        - Mean=0.7063 with IQR of 0.124
    - Can Change Case Manager:
        - 37 States with 15,330 patients
        - Mean=0.8773 with IQR of 0.086
    - Person-Centered Goals
        - 37 States with 6,716 patients
        - Mean=0.9103 with IQR of 0.073
    - ADL Goal:
        - 35 States with 2,690 patients
        - Mean=0.7749 with IQR of 0.177
    - Lifelong Learning:
        - 37 States with 11,977 patients
        - Mean=0.1109 with IQR of 0.071
    - Can Stay Home When Others Leave:
        - 37 States with 16,215 patients
        - Mean=0.4525 with IQR of 0.171
    - Life Decisions Scale:
        - 37 States with 13,727 patients
        - Mean=0.6380 with IQR of 0.116
    - Respect for Personal Space
        - 37 States with 12,345 patients
        - Mean=0.9018 with IQR of 0.034

- Transportation Availability:
    - 37 States with 11,925 patients
    - Mean=0.8777 with IQR of 0.068
- Community Inclusion:
    - 37 States with 18,026 patients
    - Mean=0.8596 with IQR of 0.042
- Satisfaction with Community Inclusion:
    - 37 States with 17,532
    - Mean=0.5940 with IQR of 0.117

Disparities

- The developer reports disparities data by residence setting and race/ethnicity for the 14 individual measures. Part of the results are provided below. Please see "Table 1b.4" of the "SuppTable" for the full compilation of data.
- Data consistently indicates that performance for nearly all 14 of the measures and collectively is the highest among residents who own their home or apartment and is lowest among residents living in ICF/IID, nursing facility or other institutional setting. Due to size limitations, the corresponding table for *Measure Score by Residence Setting* data is not available below. Please see the externally linked file titled "SuppTable" and navigate to Table 1b.4a.

Measure Score by Race and Ethnicity:

Measure Name	Non-Hispanic White	Non-Hispanic Black	Hispanic or Latinx	Non-Hispanic Asian or Pacific islander	Other Race
Community Job Goal	0.359	▼0.338	0.382	▼0.338	▲0.410
Person-Centered Goals	0.911	▲0.926	0.877	0.922	0.886
ADL Goal	0.780	0.800	▼0.766	▲0.830	▼0.766
Lifelong Learning	0.776	0.803	0.800	▲0.859	▼0.753
Satisfaction with Community Inclusion Scale	▲0.612	▼0.529	0.580	0.583	0.575
Social Connectedness	0.889	0.878	▼0.865	▲0.923	0.878
Has Friends	▲0.800	0.797	▼0.749	0.778	0.794
Transportation Availability Scale	▲0.881	0.877	▼0.849	0.874	0.862
Community Inclusion Scale	▲0.858	0.845	0.843	▼0.820	▲0.860

Measure Name	Non-Hispanic White	Non-Hispanic Black	Hispanic or Latinx	Non-Hispanic Asian or Pacific islander	Other Race
Chose Staff	▲0.703	0.692	▼0.631	0.645	0.688
Can Change Case Manager	▲0.876	0.859	0.807	▼0.770	0.862
Can Stay Home When Others Leave	▲0.424	▲0.422	0.329	▼0.300	▲0.422
Life Decisions Scale	▲0.627	0.588	0.550	▼0.528	0.618
Respect for Personal Space Scale	0.902	0.896	0.890	▼0.885	▲0.905

Note: (▼) indicates that this group's score ranks among the lowest in all groups; (▲) indicates that this group's score ranks among the highest in all groups

- The data show that performance rates are higher for majority of the 14 performance measures among the non-Hispanic white demographic. In contrast, measure performance scores are generally lower among Non-Hispanic blacks, Hispanic or latinx, and non-Hispanic Asian or pacific islander, in that order.

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?
- Does the disparities data demonstrate enough distinction to compel further observation and specific intervention?

Preliminary rating for opportunity for improvement: ☐ High ☒ Moderate ☐ Low ☐ Insufficient

Committee Pre-evaluation Comments:
Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus: For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure.

- Moderate
- The study conducted by the University of Minnesota Researchers which engaged stakeholder groups across the country in Participant Planning and Decision-making processes yielded results indicating a high degree of stakeholder support. In addition, these stakeholder groups recommended revisions to the measurers related to transportation, employment and self-determination. Such feedback provides evidence that stakeholders valued and found the measures to be meaningful.
- Evidence is there, but the evidence varies because this measure is really 14 separate measures and the evidence is different for each. Still, the overarching evidence for the group as a whole seems good.
- Evidence shows the value of this measure to target population. Does support measure focus
- Based on the past 20 years of effort and the direction for “agencies to improve accountability, more effectively track outcomes, and strengthen their capacity to assess and improve service quality and responsiveness” there are two justifications for the measure. The first is past performance with patients and the trend to “support for individuals with developmental disabilities and their families.” The second is organizational performance at facilities. Based on the past 20 years of observation, we do not need at this point to re-examine the need. The evidence exists.
- I did not see actual data from patients that support this measure concept - however, the measure intent is consistent with patient and family feedback I have heard as a developer. I think the measure concept is valuable, but feel that there could be more and clearer information about patients' priorities supporting the domains and individual items.

1b. Performance Gap: Was current performance data on the measure provided? How does it demonstrate a gap in care (variability or overall less than optimal performance) to warrant a national performance measure? Disparities: Was data on the measure by population subgroups provided? How does it demonstrate disparities in the care?

- Moderate gap in care
- Not all states participated in the study, so there is opportunity for improvement in state participation rates. However, for performance data that reflected more than 74% state participation, the data supports opportunities for improvements in several measures (e.g., community job, life decisions, community inclusion, and can stay at home when others leave). When performance measures are compared across racial and ethnic groups non-Hispanic whites tend to fare better than all other non-Hispanic and Hispanic population groups.
- Gaps are different for each of the 14 measures. Some are substantial and others seem much less so.
- Data was provided that does show a need for a national performance measure. Disparities data shows lower scores among marginalized groups.

- Yes, based on table 10 there is inconsistent care across states with 11 or almost 30% of states performing below benchmark to ability to stay home alone and satisfaction with community inclusion. Based on the data, typically 20% of the measures have states performing below benchmark. We need to do further understanding of Hispanic or Latinx, non-Hispanic Black and non-Hispanic Asian measures. The sample size is not reflective of the total population mix. Going forward this needs to be addressed in the roll-out of testing. It is non-productive to produce results that it is already known that the comments will be that the sample sizes are skewed to one population. It creates re-work and diminishes the value of the results because patient populations are not proportioned to reflect use.
- Disparities documented; measure performance gap documented - seems to meet this criterion

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

2b. Validity: Testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability; Missing Data

Reliability

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

Validity

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

2b2-2b6. Potential threats to validity should be assessed/addressed.

Composite measures only:

2d. Empirical analysis to support composite construction. Empirical analysis should demonstrate that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct.

Complex measure evaluated by Scientific Methods Panel? ☒☐ **Yes** ☐ **No**

Evaluators: Scientific Methods Panel Subgroup 1

Full Analyses:

Scientific Methods Panel Review (Combined)
Developer Responses to SMP Review

Methods Panel Evaluation Summary:

This measure was reviewed by the Scientific Methods Panel and discussed during their evaluation web meeting. A summary of the measure and the Panel discussion is provided below:

- Reliability: H-3; M-3; L-2; I-1 (Pass)
- Validity: H-0; M-4; L-0; I-3 (Consensus Not Reached)

The measure passed on reliability but did not pass on validity during the review prior to the March 2021 SMP meeting. The focus of the discussion was therefore on validity. In their preliminary analyses, the SMP noted that the submission was not complete in the data element validity testing as the developer has only listed references to studies without appropriately summarizing their results, hence there was no data element validity evaluation conducted by the SMP reviewers. It was noted that none of the risk factors for this risk-adjusted measure were tested. Furthermore, the SMP noted that at the developer's testing of performance score validity at the state level was not optimal because all of the constructs are estimated based on the same survey, suggesting that any validity issues that affect the entire survey in a consistent manner are likely to lead to exaggerated correlations. The SMP suggested analyses with external measures of quality in a comparable quality domain would have been more appropriate. The SMP further noted that the results of the Pearson product-moment correlation analysis are difficult to interpret since the theoretical relationship to the correlates chosen was not provided, nor whether only significant associations were returned/presented.

The developer provided responses to each of these critiques to the SMP:

- **Issue 1:** Confirmatory factor analysis results were not reported for multi-item scales. Developer responded by reporting results of a confirmatory factor analysis evaluating the factor structure of the five multi-item measures with results indicating that the data fits well.
- **Issue 2:** Using correlations to support validity. Developer responded by expanding their presented analysis to include external measures of quality (not just between the 14 survey items) with results that were directionally appropriate, of statistical significance and moderate to high strength in the association.
- **Issue 3:** There were concerns regarding whether or not the submitted measures should be considered PRO-PMs. The developer pointed out that NQF's definition from *Patient Reported Outcomes (PROs) in Performance Measurement* (NQF, 2013):

  > "Patient-reported outcomes (PROs) are defined as 'any report of the status of a patient's (or person's) health condition, health behavior, or experience with healthcare that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else.'"

  The measure of course meets this definition.
- **Issue 4:** Panel members sought additional evidence of validity through connection to other measures of quality. Developer responded by noting that they did not submit crosswalk testing between NCI and other measures, but that these analyses have previously been conducted. Specifically, the developer provides results to state level studies in Arizona, Massachusetts, Kentucky, Tennessee and Michigan. The NQF staff would also refer to the response in Issue 2.
- **Issue 5:** Exclusion criteria were unclear and inconsistent on MIF and testing forms. One panel member noted that "…the MIF notes a number of exclusions. However, the testing form checks the box for 'no exclusions'…" The developer provides a detailed explanation of the division of the survey and the nature of exclusions within the measure.

- **Issue 6:** States cherry-pick favorable surveys or survey sites. The developer notes that survey strategies in the states are designed by third parties through workplans. This precludes states from picking successful sites or programs for interviewing.

During the meeting, the SMP discussed a missing theory of quality between the 14 measures under the measure heading. The SMP also pointed out the possibility of a confounding influence of other factors. By way of example, they point to the first measure related to community job goals. The argument is that because urban settings provide greater job opportunities, one would expect a correlation at the state level between the percent of people who live in urban areas and the score of the measure, and indeed there is such a correlation. The SMP expressed uncertainty that this establishes the validity of the measure, but that rather it may be a confounder that ought to be adjusted for because urbanicity is not a dimension of quality. The SMP asked the developer to describe how the pattern of relationships between each of the measures that the developer described establishes the validity of each of the measures. The SMP emphasized that the submission would benefit from a clear explanation of the quality construct for the measure. The developer noted that they had provided information related to this in the responses to SMP concerns, specifically responses that (1) provide theoretical/hypothetical context for the reported Pearson correlation coefficients, (2) correlate measures with external data, (3) report complete correlation results with proper corrections and (4) provide information about #PCP-1 (Community Job Goal), #PCP-3 (ADL Goal), #CI-1 (Social Connectedness), and #PCI-3 (Transportation Availability Scale). The developer articulated directional hypotheses for expected associations among measures and only tested those hypotheses, noting that all 14 measures were supported in at least one hypothesis. The developer provided a table that summarized the results within their response. In the discussion, the SMP noted that the provided table addresses the heart of the concerns. The developer further noted that the home and community-based services (HCBS) report developed by NQF in 2016 provides a theoretical framework for HCBS quality and expressed that the measure itself is aligned with that quality framework. As an example, the developer suggested that a high-quality HCBS might provide a person-driven system to optimize individual choice, which aligns with the Choice and Control quality domain described in the 2016 HCBS report.

The SMP further pointed out that the calibration results for two of the measures in the case mix adjustment contained large discrepancies between observed and predicted values across deciles and predictive risk. The developer noted that a deeper understanding of this would require additional analysis, but the life decision scale in particular is made of several factors that are actually stand-alone instrument items and noted a tradeoff between more inclusion or better availability of the score. The SMP also expressed concern about variability between interviewers across states. The developer noted that there are training regimens and train the trainer programs that aim to reduce variability between interviewers administering the survey.

Ratings for reliability: H-3; M-3; L-2; I-1; → Measure passes with MODERATE rating

- Reliability testing conducted at the data element level:
    - The developer describes multiple data element analyses conducted, some from previous work conducted and others based on a relatively recent sample of In-Person Surveys (IPS) of the National Core Indicators (NCI). Sample includes 37 states and a total of 22,000 completed surveys.
    - IRR studies summarized by percent agreement and Kappa statistics. Studies conducted in 1997, 1998, 1999, 2008, and 2010 resulted in the following average agreements/kappa scores:
        - 1997: average agreement of 93 percent;
        - 1998: average agreement of 93 percent/average kappa score of 0.79;
        - 1999: average agreement of 92 percent;

        - 2008: average kappa score of 0.90;
        - 2010: 80 percent/average kappa score of 0.89.
    - Cognitive tests for data elements were conducted to ensure respondent understanding. The number of valid responses to the items ranged between eight and 10 with an overall average of nine over all 35 items.
    - Principal components analysis (PCA) exploratory factor analysis: The five items comprising the Life Decisions Scale constituted a single factor that explained 45 percent of the variance with component loadings ranging from 0.521 to 0.759.
        - The three items comprising the Respect for Personal Space Scale constituted a single factor that explained 44 percent of the variance with component loadings ranging from 0.490 to 0.746.
        - The two items comprising the Transportation Availability Scale constituted a single factor that explained 71 percent of the variance with both components loading at 0.842.
        - The four items comprising the Community Inclusion Scale constituted a single factor that explained 44 percent of the variance with component loadings ranging from 0.583 to 0.720.
        - The five items comprising the Satisfaction with Community Inclusion Scale constituted a single factor that explained 46 percent of the variance with component loadings ranging from 0.583 to 0.763.
    - Internal consistency of scales
        - Cronbach's alpha for multi-item scales
            - Life Decisions Scale: 0.686
            - Respect for Personal Space Scale: 0.349
            - Community Inclusion Scale: 0.687
            - Satisfaction with Community Inclusion Scale: 0.704
        - Spearman-Brown coefficient for two-item scales: The Transportation Availability Scale was 0.591.
        - Corrected Item: Total Correlation Coefficients ranged from 0.134-0.551 across the 14 items.
- Reliability testing conducted at the score level:
    - The developer conducted ANOVA to assess between-state variance in relationship to within-state variance and assessed inter-unit reliability (IUR). ANOVA analysis found that between-state variation is significantly larger than within-state variation for each of the 14 measures ($p<0.001$). The IUR ranged between 0.753 and 0.984.

Ratings for validity: H-0; M-4; L-0; I-3 → Consensus Not Reached

- Validity testing conducted at the data element level:
    - The developer suggests that interviewers are asked to give formal feedback on interviews conducted to ensure individual interview validity.
    - The developer provides seven studies investigating the relationships among NCI data elements and testing hypotheses about expected associations.
    - Developer provided responses to all SMP concerns.
- Validity testing was conducted at the score level:
    - The developer reports Pearson Product Moment Correlation Coefficients among the 37 states performance scores between the 14 IPS items.
    - The score range from 0.345-0.763 suggests a moderate to high correlation result.
    - Developer provided responses to all SMP concerns.

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The Scientific Methods Panel (SMP) is satisfied with the reliability testing for the measure. Are there additional items related to the reliability of the measure that should be discussed?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- The SMP is did not reach consensus with the validity analyses for the measure. How would the Committee rate the validity of the measure in light of the developer's written responses to the SMP's concerns?

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Preliminary rating for validity: ☐ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient** ☒ **Consensus Not Reached**

Committee Pre-evaluation Comments:
Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Reliability-Specifications: Which data elements, if any, are not clearly defined? Which codes with descriptors, if any, are not provided? Which steps, if any, in the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) are not clear? What concerns do you have about the likelihood that this measure can be consistently implemented?

- Reliable
- I do not have any concerns about the ability to consistently implement the measures.
- definitions OK
- n/a
- Reliability requires that consistent sampling is obtained. There could be variances that are not discovered because the sample size did not adequately reflect the total patient population.
- Insufficient data regarding how each item was developed/tested to determine if there are fundamental issues with reliability of specifications

2a2. Reliability - Testing: Do you have any concerns about the reliability of the measure?

- No concerns
- None noted.
- no
- n/a
- The data points measured are subjective to the person. For example, community inclusiveness is how a person feels about loneliness. This may vary by ethnic background. For this measure over 12,000 non-Hispanic white individuals were sampled yet only 430 no-Hispanic Asians. There seems to be an opportunity to cross reference by state and ethnic background. It is unclear if trends emerge identifying micro differences. Therefore the reliability to consistently implement the measure across states, demographics and ethnicity can be questioned.
- testing provided supports acceptable measure reliability

2b1. Validity -Testing: Do you have any concerns with the testing results?

- No concerns
- The developer did not provide sufficient information to inform decision-making about the measure's validity, which was noted by the comments of the Scientific Methods Panel.
- no concerns
- n/a
- In light of the written responses, I would support the measure. I believe it is more important to collect data then to over analyze and not get input that may uncover real patient issues by population.
- lack of external quality standard for empirical analyses is problematic but I applaud creative comparisons to other metrics such as state resources and licensure; measure does have face validity however

2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data) 2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure?

- No threats to validity
- I share a similar perspective of one of the SMP member's comments about the possibility of bias (states that "cherry picking") among states when choosing where to administer the survey and how many, which could be a threat to validity.
- Analyses are different between the 14 measures that are part of this whole measure, mainly significant/meaningful differences, making validity variable across the set of measures..

- n/a
- Missing data does constitute a threat to the validity of this measure because the data collection plan is heavily weighted to one population. Going forward, the data collection plan needs to address how the sample sizes will reflect the breakdown of patients by demographics and ethnicity.
- A threat to validity is moderate response rate and possible case selection/response bias could be fatal flaw it this is not addressed through either risk adjustment/stratification (not currently) or implementation

2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment) 2b2. Exclusions: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure? 2b3. Risk Adjustment: If outcome (intermediate, health, or PRO-based) or resource use performance measure: Is there a conceptual relationship between potential social risk factor variables and the measure focus? How well do social risk factor variables that were available and analyzed align with the conceptual description provided? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Was the risk adjustment (case-mix adjustment) appropriately developed and tested? Do analyses indicate acceptable results? Is an appropriate risk-adjustment strategy included in the measure?

- No other threats to validity
- The comments of the SMP raise concerns for me about the approach used for risk adjustment. Subsequent discussion of implications of missing data on results and the differences in characteristics of respondents (e.g., group home vs familial residential status) and lack of stratification by residential setting suggests potential variable influences on responses.
- no concerns
- n/a
- The risk-adjustment strategy for case-mix by ethnic breakdown needs further investigation.
- See above - and unclear why some measures are adjusted while others aren't. Raises question about whether all component measure stand on their own and whether SDOH appropriately considered through stratification or other means

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. **Feasibility** is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- The developer notes challenges with feasibility, including challenges with data collection for the 38 states collecting NCI data for ID/DD HCBS measures, but notes that most states reported that the identified challenges were overcome once initial processes and protocols were established and subsequently repeated.
- The developer also highlights challenges associated with data confidentiality/data access for states that are under contract with external administrative entities as well as sample identification challenges facing states that elect to oversample or stratify data by population.
- The developer provides details regarding how these challenges are managed to improve feasibility.

Questions for the Committee:

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?
- Is the data collection strategy ready to be put into operational use?

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:

Criteria 3: Feasibility

1. **Feasibility: Which of the required data elements are not routinely generated and used during care delivery? Which of the required data elements are not available in electronic form (e.g., EHR or other electronic sources)? What are your concerns about how the data collection strategy can be put into operational use?**

- Measure is feasible
- No concerns noted based on the developer's discussion of the initial challenges that are typical or states that involve the adoption of processes and protocols to support participation in data collection activities.
- no concerns
- Challenges are noted with data collection, would like to hear more about this from developer
- We cannot let the excuse of it is a burden to implement this measure become justification for not moving forward. While it may create addition burden of healthcare workers, the overall impact of having the data to know the patient perspective needs to be valued.
- high burden measure due to survey format but necessary to get at these complex and important concepts

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact/improvement and unintended consequences

4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

4a. Use evaluates the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No** ☐ **UNCLEAR**

OR

Planned use in an accountability program? **Yes** ☒ **No**

- Developer indicates plan to use measure in quality improvement (external benchmarking to organizations) programs
- The developer reports this measure is currently in use for:
  - Public reporting:
    - Medicaid Adult Core Health Care Quality Measure Set
      https://www.medicaid.gov/medicaid/quality-of-care/performance-measurement/adult-and-child-health-care-quality-measures/adult-health-care-quality-measures/index.html

- Regulatory and Accreditation Programs
  - Connecticut Medicaid 1915(c) HCBS Waiver Assurances

    https://portal.ct.gov/-/media/DDS/events2020/EDS_Waiver_Renewal_10_27_2020.pdf
  - Indiana Family and Social Services Administration Medicaid 1915(c) HCBS Waiver Assurances

    https://www.in.gov/fssa/dmha/files/3.1i.pdf
- Quality Improvement (Internal to the specific organization)
  - Arizona: Monitored provider rates as the state incentivized Community and Supported Employment initiatives

    https://www.aaidd.org/publications/bookstore-home/product-listing/cross-cultural-quality-of-life-enhancing-the-lives-of-people-with-intellectual-disability-second-edition
  - Massachusetts: The state Department of Developmental Services used measure of loneliness and friendships to identify areas for quality improvement in supporting relationships

    https://shriver.umassmed.edu/wp-content/uploads/2020/07/QINA-Friendship_final_web2.pdf
  - Kentucky Division of Developmental and Intellectual Disabilities reviews outcomes from National Core Indicators with stakeholders to identify recommendations for quality improvement

    http://www.kentuckycq.org/wp-content/uploads/2018/02/QI-draft-recom_2017_final.pdf

Accountability program details:

- The developer does not provide a plan for use within an accountability application.

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

- The developer states that 19 out of 22 states that provided written feedback to National Core Indicators (NCI) responded that they were satisfied or very satisfied, and three states indicated a neutral response concerning the measures and the overall data collection and reporting process.
- The developer also explains that 13 states strongly agreed, eight states agreed, and one state disagreed when asked if NCI as a good value for the state’s DD (Developmental Disability) system.
- The developer adds that the feedback and information gathered by states from the surveys provides recommendations for new domains/measures to be added and informs the content of the survey questions and the wording of survey questions.
- The developer also received feedback from users. The developer indicated that the results of the user surveys revealed that a third of the questions had a request for repetition (i.e., respondent asked the question to be repeated) and slightly over half of the questions had a clarification request from at least

one of the 10 participants. Additionally, the developer found that about 47% of the questions had no repetition or clarification request from any of the respondents.

    - The developer explains that the participants found the questions to be easy to understand and to answer and that the survey results reveal that for the majority of participants, the source of the difficulty was not cognitive burden, but instead one of emotional (e.g., the question on loneliness triggered sadness and made it difficult to answer) or situational (e.g., respondent was tired).
    - The developer notes that most of the clarification requests were for additional information that surveyors are routinely trained to provide, such as alternative wording or examples for the key concept(s).
    - The developer notes that the results suggest a few minor revisions to question wording or the response options, which will be included in planning discussions for the next revision cycle.

Additional Feedback:

- N/A

Questions for the Committee:

- How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?
- How has the measure been vetted in real-world settings by those being measured or others?

Preliminary rating for Use: ☒ **Pass** ☐ **No Pass**

4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

4b. Usability evaluates the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

Improvement results

- Developer provides the interpretation of results from one measure (#PCP-1) contained in the subset of measures that, combined, constitute the five measures in the HCBS Domain: Person-Centered Planning (PCP) and Coordination.
    - #PCP-1 measures the proportion of people who express they want a job who have a related goal in their service plan (Community Job Goal). The measure #PCP-1 is used by the Supported Employment Leadership Network (SELN) to measure each participating state's progress in utilizing supported employment services.
        - The developer mentions that the Missouri DD agency utilized this measure in combination with other NCI measures to improve rates of employment across the state. Missouri rates for people who had a paid community job increased from 9% in 2017-18 to 14% in 2018-19.
- The developer provides data that demonstrate increased state- and user-level engagement.
    - The developer reports that between 2014 and 2020, 20 additional states have joined as NCI members and the total number of individual surveys entered per year has increased from 13,157 in the 2012-13 reporting cycle to 22,009 in 2018-19.
    - The developer states that NCI has continuously increased its reach and application across state DD systems.

4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation

- The developer indicates that there were no unexpected positive or negative unintended findings.

Potential harms

- No potential harms are identified by the developer.

Additional Feedback:

- None

Questions for the Committee*:*

- How can the performance results be used to further the goal of high-quality, efficient healthcare?
- Do the benefits of the measure outweigh any potential unintended consequences?

Preliminary rating for Usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 4: Usability and Use

4a1. Use - Accountability and Transparency: How is the measure being publicly reported? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? For maintenance measures - which accountability applications is the measure being used for? For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided? 4a2. Use - Feedback on the measure: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? Has this feedback has been considered when changes are incorporated into the measure?

- Appropriate Use
- The Developer indicates the measures are being publicly reported in Medicaid and regulatory and accreditation programs. Some states engage stakeholders to review the results as part of quality improvement efforts.
- no concerns
- Minor revisions were made based on user testing.
- ? This measure is a real-world insight to the patient experience. The medical outcome of an intervention goes beyond the four walls of a facility. Healthcare is about the factors of planning, coordination, community inclusion, connectedness, transportation and other measured factors. When we measure the experience of the healthcare journey, we can achieve high-quality, efficient healthcare. Therefore, transparent reporting of these measures to the public will help determine by demographics, ethnicity, and location best practices that can be transferred to other healthcare organizations.
- in use in multiple programs

4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? If not in use for performance improvement at the time of initial endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations? 4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them.

- Measure is usable
- Greater and more consistent use of the measures as part of quality improvement efforts that support actionable efforts to enhance program performance may lead to more efficient care and better outcomes. The Developer reported no unintended consequences were experiences. Such outcomes suggest minimal harms associated with the measure. Perhaps a potential harm may relate to any instances where confidentiality breeches occur.
- no concerns
- No harms noted. Survey usage has increased over time and can be used to improve care.
- We need to look at healthcare from the total delivery perspective which this measure is designed to achieve. While measure may have reliability, validity and feasible issues to overcome, using the obstacles as justification to not implement the measure outweighs the harms over the benefit of having the data to determine the overall patient experience.
- if non-response is not appropriately handled, then misleading measure results could lead to negative consequences such as diversion of resources away from areas that have critical need

Criterion 5: Related and Competing Measures

Related or competing measures

2967: CAHPS® Home- and Community-Based Services Measures

Harmonization

- The developer identifies measure #2967 as a related but not competing measure and indicates that measure specifications are not harmonized.
    - The developer elaborates on the distinctions by explaining that NQF 2967 - CAHPS Home and Community Based Services Measures could be used to survey the same population as it is described as a cross-population survey, but that NCI for ID/DD HCBS measures are specifically designed to survey the target population of adults with intellectual or developmental disabilities who are receiving HCBS.
    - Furthermore, the developer states that the NCI for ID/DD HCBS measures do not have the same focus as HCBS-CAHPS measures and explains that one area which merits mention is the transportation item because it may appear to be related with a similar focus.
        - The Transportation availability scale includes a measure of having transportation available when needed. This is not the same measure as the “Transportation to Medical Appointments” scale that exists as part of HCBS-CAHPS, which only focuses on medical appointments.
        - The developer adds that Home and Community Based Services (HCBS) are intended to support people to live a life in the community that extends beyond merely medical appointments, therefore a measure of broader access to transportation is important to have.

Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures

5. Related and Competing: Are there any related and competing measures? If so, are any specifications that are not harmonized? Are there any additional steps needed for the measures to be harmonized?

- No
- A competing measure noted was 2967 : CAHPS® Home- and Community-Based Services Measures, however, it was noted that the NCI for ID/DD HCBS Measures were specifically designed to survey the target population of adults with intellectual or developmental disabilities who are receiving HCBS.
- Yes - 2967 is related, not competing.
- For this specific population, there does not seem to be other competing measures.
- overlaps a little with CAHPS, but CAHPS not intended for individuals with disability so not directly competing

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 06/10/2021

- No NQF Members have submitted support/non-support choices as of this date.
- No Public or NQF Member comments submitted as of this date.

Scientific Acceptability

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 3622

Measure Title: National Core Indicators for Intellectual and Developmental Disabilities (ID/DD) Home- and Community-Based Services (HCBS) Measures

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?** ☒ **Yes** ☒ **No**

   Submission document: "MIF_xxxx" document, items S.1-S.22

   ***NOTE**: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

2. **Briefly summarize any concerns about the measure specifications.**

   Panel Member 3: The sampling specifications are unclear. The application states that each state is "recommended to have a sample size that will support both a 95%confidence level and a ±5% margin of error, but not what measures those are based on, minimum recommended sample size, sampling procedures, etc. Without more detail there is concern that sampling procedures could vary substantially by state and therefore be a source of bias in between state comparisons. Also, sample sizes vary substantially by state. Minnesota and Texas account for 20% of the 37-state sample.

   Panel Member 4: In the MIF (S.7 – S.9) it is not fully specified as to the logic of excluding surveys due to responses in Section 1. Thus, the lack of specificity / guidance as to exclusions increases the likelihood of inconsistent survey scoring from surveyor to surveyor.

   Panel Member 7: Everything is clearly specified. According to NQF, "patient-reported outcomes (PROs) are defined as "any report of the status of a patient's (or person's) health condition, health behavior, or experience with healthcare that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else... the word "patient" is intended to be inclusive of all persons, including patients, families, caregivers, and consumers more broadly. It is intended as well to cover all persons receiving support services, such as those with disabilities. Key PRO domains include: Health-related quality of life (including functional status); Symptoms and symptom burden (e.g., pain, fatigue); Experience with care; and Health behaviors (e.g., smoking, diet, exercise)." My only question or concern is that some of the items seem to fall outside traditional PRO and hence PRO-PM boundaries; for example, "The proportion of people who report adequate transportation." Given that ID/DD agencies are responsible for providing and managing a comprehensive range of social and support services, are there any boundaries to the subset of services that NQF-endorsed measures can evaluate? Are transportation services "fair game," to the extent that available and affordable transportation is necessary to optimize quality of life? But this is true for all persons, not just those with ID/DD! Clearly educational services promote healthy behaviors and self-fulfillment for all persons, not just those with ID/DD... so are all measures evaluating educational services, for all subpopulations, also within NQF's domain? When they draft their Importance section, the developers should clearly articulate the conceptual foundation for treating all 14 of these measures as PRO-PMs (or identify the subset that they are advocating to be regarded as PRO-PMs).

Panel Member 9: None

RELIABILITY: TESTING

Type of measure:

☒ Outcome (including PRO-PM) ☐ Intermediate Clinical Outcome ☒ Process

☐ Structure ☐ Composite ☐ Cost/Resource Use ☐ Efficiency

Data Source:

☐ Abstracted from Paper Records ☐ Claims ☐ Registry
☐ Abstracted from Electronic Health Record (EHR) ☐ eMeasure (HQMF) implemented in EHRs
☒ Instrument-Based Data ☐ Enrollment Data ☒ Other (please specify)
Panel Member 1: National Core Indicators® In-Person Survey (IPS) Data
Panel Member 2: Survey
Panel Member 4: National Core Indicators In=Person Survey data
Panel Member 5: National Core Indicators® In-Person Survey (IPS) Data
Panel Member 6: Interview with patient
Panel Member 8: "administrative data" is referred to for Background Information, but it is not exactly clear which administrative data sets are used

Level of Analysis:

☐ Individual Clinician ☐ Group/Practice ☐ Hospital/Facility/Agency ☐ Health Plan
☒ Population: Regional, State, Community, County or City ☐ Accountable Care Organization
☐ Integrated Delivery System ☒ Other (please specify)
Panel Member 1: State
Panel Member 4: state
Panel Member 5: State
Panel Member 9: state

Measure is:

☒ New ☐ Previously endorsed (NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)

Submission document: “MIF_xxxx” document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level ☒ Measure score ☒ Data element ☐ Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure**
   ☒ Yes ☐ No
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of **patient-level data** conducted?

   ☐ Yes ☐ No
6. **Assess the method(s) used for reliability testing**

Submission document: Testing attachment, section 2a2.2

Panel Member 1: IUR

Panel Member 2: The developers performed detailed and complete analyses of reliability both for data elements and overall scores.

Panel Member 3: The developer used Cohen's kappa to assess inter-rater reliability based on either in-person or videos of interviews with respondents. Data are provided for 20 interviews rated by 6 trained surveyors from a 2010 study. Cronbach's alpha was performed for multi-item scales, with Spearman Brown coefficients used for reliability assessment of the Transportation Availability Scale. Item-to-total correlations are provided for the collective groups of 5 multi-item scales, but it is not clear what analysis produced the results. Was an internal consistency reliability analysis performed for all items as a group (possibly evaluating a global underlying construct) or for each scale separately. Factor analyses are typically performed to assess convergent validity and will be treated as evidence of validity for this review: Otherwise, reliability analyses performed appear appropriate.

Panel Member 4: Testing methods for data element and measure score were appropriate for the 14 measures that comprise 3622. Critical data element testing: Regarding the in-person survey: Inter-rater reliability studies where the results were summarized by calculating percent agreement among raters and Cronbach's Kappa. Regarding data collected from administrative records: The data were tested with a number of individuals abstracting the information from the records. Regarding critical data elements: Cognitive testing was conducted with people from the target population. The test identifies how well the survey questions were understood by the respondent. Measure score testing: ANOVA test was used with the 14 measures in regard to whether the between state variance is larger than the within state variance. Additionally, an inter-unit reliability (IUR) test was performed for each measure, which is essentially a SNR test.

Panel Member 5: Reliability Tests on Critical Data Elements was assessed using -inter-rater reliability (IRR) studies measured using percent agreement among raters and Cronbach's Kappa: kappa score mean is 0.89 indicating high reliability -Reliability Tests on Performance Measure Scores was measured using the IUR - IUR ranged between 0.75 and 0.98 for each of the 14 measures, which is consistent with high-reliability

Panel Member 6: All are ok for 1st submission.

Panel Member 7: For data element reliability, developers assessed inter-rater reliability (estimated as percent agreement and the kappa coefficient), cognitive testing, and Cronbach's alpha for internal consistency of multi-item scales (Spearman-Brown for 2-item scales). For performance score reliability, they assessed inter-unit reliability based on ANOVA. All of these testing procedures are appropriate, but the developers properly noted that IUR is exaggerated by the role of state legislation and regulation, as well as unmeasured differences in case mix (eligibility) across states. I wish it were possible to factor out the portion of state-level signal that is due to the ID/DD service provider (but it isn't).

Panel Member 8: inter-observer reliability kappa and Cronbach kappa and Spearman Brown for data elements IUR among states for measure also, cognitive test

7. **Assess the results of reliability testing**

Submission document: Testing attachment, section 2a2.3

Panel Member 1: IUR >0.7 for all measures

Panel Member 3: Inter-rater reliability results are based on the 2010 study show Kappa statistics averaged .89 (range: .82-.95), indicating good reliability. Cronbach's alpha for multi-item scales approached .70 for 3 of the 5 scales, but was, .35 for the respect for personal space scale; the Spearman Brown coefficient was .59 for the transportation availability scale. IUR results reported ranged from .75-.98, within guidance for acceptable between state differences vs. variation within states. Inter-rater reliability and IUR results are within guidance for high reliability.

Panel Member 4: Regarding data element testing: Issues with the sections / domains with test results: [1] They do not always match up with the 4 domains of 3622. Example: The MIF states a section is titled “Choice and Control”. However, on p. 10 of the testing form, the 2 sets of results on top of the page do not include results for the “Choice…” section. [2] They differ from various test results reported. Example: Page 10 has test results for section titled “Respect for Personal Space”. However, test results on p. 9 do not report test results for this section. [3] The test results bottom p. 9 that have a crosswalk between section name & measure ID, these section names do not carry through in subsequent test results. Example: Page 10 has test results for section titled “Transportation Availability”, but that section doesn’t appear in the aforementioned p. 9 test results. Thus, we don’t know what measure IDs within “Transportation”. Most measures demonstrate moderate reliability. The exceptions follow: PCP-5: high; CI-4: high; CC-3: low Regarding measure score testing: Given the IUR results spanned 0.753 to 0.984, I would rate the measures as “moderate” given the response provided does not report the 14 individual measure scores.

Panel Member 6: Some of the results are concerning such as the Kappa scores for the home section, the Cronbach's alpha for the Respect for Personal Space scale and the errand question.

Panel Member 7: Inter-rater reliability is generally excellent, although stronger for the multi-item scales (PCP-5, CI-4, CC-4, HLR-1) than for some of the single items (CC-3, CI-1, CI-2, PCP-2). Have the developers considered combining some of these items, such as CI-1 and CI-2, and PCP-1 and PCP-2, to create more robust and reliable scales? My general impression is that the "reliability story" would be enhanced by having somewhat fewer measures (e.g., 8-10) with at least 2-3 items in most measures. On the other hand, the Respect for Personal Space Scale has poor internal consistency reliability, apparently because having "a place to be alone in your home" does not correlate with advance notification before entry.

Panel Member 8: All data element tests were conducted on small within state samples--how states were chosen and how participants were chosen is obscure; therefore, meaning of results is difficult to interpret or generalize.

Panel Member 9: Data element-level tests report adequate results, although results are complicated given 14 different measures included here. Measure score reliability in analyses at the state level seems strong for most of the 14 measures.

8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

   ☐ **Not applicable** (score-level testing was not performed)

9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

   ☒ **Not applicable** (data element testing was not performed)

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and **all** testing results):

    ☒ **High** (NOTE: Can be HIGH **only if** score-level testing has been conducted)

    ☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has **not** been conducted)

    ☒ **Low** (NOTE: Should rate **LOW** if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

☒ **Insufficient** (NOTE: Should rate **INSUFFICIENT** if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

    Panel Member 2: Score-level estimates of inter-unit reliability were above 0.75 for all 14 individual measures and were above 0.90 for 9 of them. These high score-level reliability results were obtained despite having mixed results for reliability at the item level.

    Panel Member 3: Respondent level reliability results were lower than acceptable for group comparisons. Therefore, the reliability would be rated as moderate.

    Panel Member 4: In the MIF (S.7 – S.9) it is not fully specified as to the logic of excluding surveys due to responses in Section 1. Thus, the lack of specificity / guidance as to exclusions increases the likelihood of inconsistent survey scoring from surveyor to surveyor. [response to Q2] Given the instruction by "low" states "…rate LOW if you believe specifications are NOT precise, unambiguous…", the rating here is "low".

    Panel Member 5: Reliability Tests on Critical Data Elements was assessed using -inter-rater reliability (IRR) studies measured using percent agreement among raters and Cronbach's Kappa: kappa score mean is 0.89 indicating high reliability -Reliability Tests on Performance Measure Scores was measured using the IUR - IUR ranged between 0.75 and 0.98 for each of the 14 measures, which is consistent with high-reliability.

    Panel Member 6: Ok for 1st submission.

    Panel Member 7: I would give a Low rating to the Respect for Personal Space Scale due to its poor Cronbach's alpha (0.349) and CITCs (as low as 0.134). The other multi-item scales would get moderate ratings for internal consistency reliability. My overall rating is insufficient because we are lacking: 1. Spearman-Brown coefficients or other measures of internal consistency for PCP-1 and PCP-3. 2. IRRs (kappa scores) at the individual measure level. We are given only percent agreement at the survey section level, not kappa scores at the measure level. Some previous studies are cited to address kappa scores for IRR, but these studies appear to have reported results at the entire-survey level rather than the measure level. By definition, all measures for NQF review must have measure-specific results.

    Panel Member 8: Tests are administered by surveyor who can certainly subjectively impact results and who judges the validity of the answers Participation is voluntary and nearly have of those identified do not participate States determine how many people to survey and where within the state surveys are conducted Tests of data element reliability are conducted with small samples within one or two states.

    Panel Member 9: As was the case for the CAHPS set of measures, this single evaluation is difficult because there are 14 distinct measures in this package. Most of them are adequate for reliability, there are a couple of shaky Cronbach's alpha scores for data element reliability, but all of the measure score reliabilities are apparently strong enough.

VALIDITY: TESTING

12. **Validity testing level:** ☒ **Measure score** ☒ **Data element** ☒ **Both**
13. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?**

    NOTE that data element validation from the literature is acceptable.

    Submission document*: Testing attachment, section 2b1.*

    ☒ **Yes**

    ☒ **No**

    ☒ **Not applicable** (data element testing was not performed)

14. **Method of establishing validity of the measure score:**

    ☒ **Face validity**

    ☒ **Empirical validity testing of the measure score**

☐ **N/A (score-level testing not conducted)**

15. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

    Submission document: Testing attachment, section 2b1.

    ☒ **Yes**

    ☒ **No**

    ☐ **Not applicable** (score-level testing was not performed)

16. **Assess the method(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.2

    Panel Member 1: Correlations among measures

    Panel Member 3: Principal components analysis was performed on individual multi-item scales and results presented appear adequate, however statistics appropriate model (e.g., CFI, RMSEA, etc.) were not provided, nor were the actual factor analysis. Further, more helpful to demonstrate convergent/discriminant validity evidence would have been a rotated factor analysis (e.g., varimax or obliques) of all multi-item scales together. The results of the Pearson product-moment correlation analysis are difficult to interpret since the theoretical relationship to the correlates chosen was not provided, whether only significant associations were returned/presented, etc. Also, Bonferroni correction for significance should have been done, particularly if a balanced correlation across all validation variables was performed and only significant ($p<0.5$) values retained.

    Panel Member 4: Regarding testing of critical data elements: Given the measure is risk adjusted, I would consider the risk factors as critical data elements. In this regard, none of the risk factors were tested. Regarding measure score testing: There is sparse reference to the fact testing was conducted regarding the relationship between 3622 and other measures. In the question regarding testing methods, it's unstated regarding: [a] the methods tested for such a relationship; [b] the tested employed for identifying such relationships; [c] hypothesized correlations and directionality of the relationships.

    Panel Member 5: examined intercorrelation among the states' performance scores showing that the NCI data elements correlate with each other.

    Panel Member 6: I did not see a theoretical framework nor comparisons to other measures - it seemed to be comparing the measure with itself.

    Panel Member 7: The developers do not provide clear evidence regarding data element validity, so this evaluation is based on measure score validity. They apparently did not do formal testing of content validity, or else they declined to report the results. They describe assessments of "the individual's comprehension of the questions and consistency of responses," but they declined to provide results of these assessments. They cite to "multiple published studies (that) investigated relationships among NCI data elements," but they do not summarize any key findings from these studies. Their testing of performance score reliability at the state level is reasonable, but not optimal, because all of the constructs are estimated based on the same survey. Therefore, any validity issues that affect the entire survey in a consistent manner are likely to lead to exaggerated correlations. For construct validity testing, external measures collected in a different way (e.g., background information collected from administrative records, surveys of ID/DD agency staff or contractors) are preferred.

    Panel Member 8: Test the correlation of the measure scores with one another. This seems problematic, as there is no external validation. Although it is reasonable to say that one might expect one measure to correlate with another, if one is not valid then correlation would suggest that none are valid.

    Panel Member 9: The developer examined correlations among the 14 measures to establish the validity of each of them. This is a weak approach to establishing validity of these as quality measures, as there is no objective measure of quality of care, linked to some conceptual definition of quality of care, that would serve as some kind of more objective grounding of validity analyses.

17. **Assess the results(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.3

    Panel Member 1: Many positive and statistically significant correlations exist.

    Panel Member 2: Some results presented in the reliability section are also relevant to the assessment of validity. Some relatively low item-level reliability results for certain measures could be a concern for validity because between-state differences could be explained in part by different interviewers.

    Panel Member 3: Principal components analysis was performed on individual multi-item scales and results presented appear adequate, however statistics appropriate model (e.g., CFI, RMSEA, etc.) were not provided, nor were the actual factor analysis. Further, more helpful to demonstrate convergent/discriminant validity evidence would have been a rotated factor analysis (e.g., varimax or obliques) of all multi-item scales together. The results of the Pearson product-moment correlation analysis are difficult to interpret since the theoretical relationship to the correlates chosen was not provided, whether only significant associations were returned/presented, etc. Also, Bonferroni correction for significance should have been done, particularly if a balanced correlation across all validation variables was performed and only significant ($p<0.5$) values retained.

    Panel Member 4: Regarding testing of critical data elements: No testing results provided. Also, as noted above in regard to the testing of critical data elements: In the questions that discusses the testing, there’s no mention of testing the risk factors. Regarding measure score testing: There was generally a moderate correlation with 3622 and the measures included in the testing.

    Panel Member 5: examined intercorrelation among the states' performance scores showing that the NCI data elements correlate with each other.

    Panel Member 6: I was expecting to see more than what was provided.

    Panel Member 7: The intercorrelations shown in Exhibit 3 are sufficient to support score-level validity, but only for those scores actually included in the Exhibit. I don't see any information on PCP-1 (Community Job Goal), PCP-3 (ADL Goal), CI-1 (Social Connectedness), CI-3 (Transportation Availability Scale). If the developers do not wish to support the validity of these four measures, they should be removed from NQF review. No information on data element validity is actually presented.

    Panel Member 8: See answer to #16

    Panel Member 9: The correlations reported are statistically significant and in the expected directions, but per note above, the whole set of correlations is not persuasive on the validity of any of the measures as quality-of-care measures. Responses to some of the measures can easily be influenced by environmental factors (economy, social networks, community cohesion) that are external to the support programs being measured.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

18. **Please describe any concerns you have with measure exclusions.**

    Submission document: Testing attachment, section 2b2.

    Panel Member 4: The MIF notes a number of exclusions. However, the testing form checks the box for “no exclusions”. In turn, there’s no response the exclusion related questions. So, we have insufficient information to assess the extent of the significance of the exclusions.

    Panel Member 7: No concerns

    Panel Member 8: As noted previously, participation is voluntary. Answers of "unknown" or "not applicable" are dropped from the denominator. Nearly 50% of planned participants do not participate.

States determine where surveys are given and how many are given (and can therefore "cherry pick" sites where they have been more successful than those where they have been less successful.

Panel Member 9: None

19. **Risk Adjustment**

Submission Document: Testing attachment, section 2b3

19a. **Risk-adjustment method** ☒ **None** ☒ **Statistical model** ☒ **Stratification**

19b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☒ Yes ☐ No ☒ Not applicable

19c. **Social risk adjustment:**

19c.1 Are social risk factors included in risk model? ☒ Yes ☒ No ☒ Not applicable

19c.2 Conceptual rationale for social risk factors included? ☒ Yes ☒ No

19c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☒ Yes ☐ No

19d. **Risk adjustment summary:**

19d.1 All of the risk-adjustment variables present at the start of care? ☒ Yes ☐ No

19d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☒ Yes ☐ No

19d.3 Is the risk adjustment approach appropriately developed and assessed? ☒ Yes ☐ No

19d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☒ Yes ☒ No

19d.5. Appropriate risk-adjustment strategy included in the measure? ☒ Yes ☐ No

19e**. Assess the risk-adjustment approach**

Panel Member 1: Basically, an additive offset for each state, based on its risk-adjusted predicted measured score.

Panel Member 2: I am confused about the calibration results in Tab 2b3.8 of the Excel document. Results appear to indicate large differences between observed versus model-predicted mean scores.

Panel Member 3: Only 2 of the 14 measures, (Life Decisions and Community Inclusion) were risk adjusted, based on conceptual information (adjustment would mask the quality-of-care differences).

Panel Member 4: It notes the 'life decision scale' & 'community inclusion scale' are risk adjusted. However, in the list of 14 measures, there is not set of measures noted as 'life decision scale'. There is a lack of a conceptual discussion as to how social risk factors were identified. The R-squared for the 'life decision scale' was 0.159 & for 'community inclusion scale' it was 0.085. In each case the findings note that the risk model is inadequate for risk adjusting these measures.

Panel Member 5: Squared values were 0.159 for Life Decisions Scale(#CC-4) and 0.085 for Community Inclusion Scale(#CI-4). Calibration, as assessed using the ratio of the observed and predicted scores by decile is acceptable.

Panel Member 6: Seems ok for the 1st submission.

Panel Member 7: Risk-adjustment is offered for the Life Decisions scale and the Community Inclusions scale. Interestingly, these two scales also have such high IURs (>0.97) that responses seem to be driven by state-level policy, resource, or case mix differences. The developers clearly articulate a rationale for risk-adjustment of these two scales but not the other 12 measures. However, they do not explain why these 12 measures are not suitable for risk stratification by residential setting. For example, it seems self-evident and unavoidable that persons living in institutional settings would have less privacy than persons living in home settings. Although state ID/DD agencies presumably have some control over residential placement, the developers have not articulated a clear conceptual basis for stratifying two of their measures by residential setting, but not others that seem just as inextricably linked to that setting.

Panel Member 8: Risk adjustment for only two of the scales--not exactly clear (to me) why these two and not the others. Limitations of risk adjustment for this type of measure are well-stated.

Panel Member 9: Since these are process measures, it is reasonable to not apply risk adjustment. One would expect that these things would be done regardless of presence/absence of clinical or social risk factors.

20. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

Submission document: Testing attachment, section 2b4.

Panel Member 3: The data provided for between state differences from global benchmarks for each measure indicate considerable state-level variation for each measure. Of concern is that less than 50% of states perform of the benchmark for 7 of the 14 indicators. If the benchmarks were created as an average across states as suggested in the submission, that finding is confusing.

Panel Member 4: There are no concerns.

Panel Member 6: None

Panel Member 7: none

Panel Member 8: Concerns regarding reliability and validity dominate my concerns rather than concerns regarding risk-adjustment.

Panel Member 9: No way to tell what difference would be meaningful.

21. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**

Submission document: Testing attachment, section 2b5.

Panel Member 2: Each state conducts their own survey and presumably employs different interviewers. Results in Section 2a2.3 and elsewhere indicate imperfect agreement between raters for some of the main survey outcome variables at the patient level. This raises the question of whether between state differences could be partly explained by different interviewers. States also differ in their response rates and rates of valid responses. Empirical analyses indicate that responders and non-responders differ with respect to their survey outcomes. I wondered if the developers considered using a stratification or weighting adjustment to account for non-response.

Panel Member 4: There was 1 set of specifications used.

Panel Member 7: not applicable

Panel Member 9: No concerns.

22. **Please describe any concerns you have regarding missing data.**

Submission document: Testing attachment, section 2b6.

Panel Member 3: The response rate appears to be slightly greater than 50% for the 2 states for which data are provided. There were significant differences in characteristics between responders and non-responders, raising concerns regarding the lack of risk adjustment for the majority of measures.

Panel Member 4: There are no concerns.

Panel Member 6: None

Panel Member 7: Nonresponse bias is well characterized based on data from two states. Based on these data, and particularly the greater tendency of responders to live in group homes rather than with parents or relatives, weighting for nonresponse (using known characteristics of the sampled individual, such as residential setting) should be considered.

Panel Member 8: As noted, there are huge gaps in data with participation voluntary and beyond that only 55% or so of patients planned to be surveyed actually completing. Dropping from the denominator any value with N/A or unknown further potentially biases the survey.

Panel Member 9: None

For cost/resource use measures ONLY:

23. **Are the specifications in alignment with the stated measure intent?**

☐ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**

24. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

25. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

☐ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

☒ **Low** (NOTE: Should rate LOW if you believe that there **are** threats to validity and/or relevant threats to validity were **not assessed OR** if testing methods/results are not adequate)

☒ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level **is required**; if not conducted, should rate as INSUFFICIENT.)

26. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

Panel Member 2: See comments above

Panel Member 3: Principal components analysis was performed on individual multi-item scales and results presented appear adequate, however statistics appropriate model (e.g., CFI, RMSEA, etc.) were not provided, nor were the actual factor analysis. Further, more helpful to demonstrate convergent/discriminant validity evidence would have been a rotated factor analysis (e.g., varimax or obliques) of all multi-item scales together. The results of the Pearson product-moment correlation analysis are difficult to interpret since the theoretical relationship to the correlates chosen was not provided, whether only significant associations were returned/presented, etc. Also, Bonferroni correction for significance should have been done, particularly if a balanced correlation across all validation variables was performed and only significant ($p<0.5$) values retained. The information provided lacks sufficient detail to evaluate the validity of these measures appropriately.

Panel Member 4: Regarding testing of critical data elements: No testing results provided. Also, as noted above in regard to the testing of critical data elements: In the questions that discusses the testing, there's no mention of testing the risk factors. Regarding measure score testing: There was generally a moderate correlation with 3622 and the measures included in the testing.

Panel Member 5: results of empirical validity testing and risk adjustment model testing are acceptable.

Panel Member 9: I don't think we can accept simple inter-correlations among sets of measures from the same survey to be evidence of validity of the set of measures as quality-of-care measures. There has to be some connection to other, preferably objective, measures of quality in order to validate the survey measures.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

27. **What is the level of certainty or confidence that the empirical analysis demonstrates that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct?**

    ☐ **High**

    ☒ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

28. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**

ADDITIONAL RECOMMENDATIONS

29. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

    Panel Member 7: Some of these 14 "measures" do not stand on their own as separate PRO-PMs. The developers should be encouraged to identify the subset that have sufficient conceptual foundation, empirical evidence of reliability, and empirical evidence of validity to support NQF endorsement as PRO-PMs.

    Panel Member 9: In spite of the developer's labeling, as in the case of CAHPS, these are not outcome measures. It is even clearer here for many of the measures than it is for CAHPS. These are things that are done or not done - processes of care.

Developer Response to SMP Preliminary Analysis

Measure Number: 3622

Measure Title: National Core Indicators for Intellectual and Developmental Disabilities (ID/DD) Home- and Community-Based Services (HCBS) Measures

Measure Developer/Steward: Human Services Research Institute, The National Association of State Directors of Developmental Disabilities Services (NASDDDS)

Validity

- **Issue 1:** Confirmatory factor analysis results were not reported for multi-item scales.
    - **Developer Response 1:** We thank the panel members for noting the omission. We conducted confirmatory factor analysis to test the factor structure of the five multi-item measures:
            - Community Inclusion Scale (CI-4)
            - Satisfaction with Community Inclusion Scale (PCP-5)
            - Transportation Availability Scale (CI-3)
            - Life Decisions Scale (CC-4)
            - Respect for Personal Space Scale (HLR-1)
        - Results indicate that the estimated model fits the data reasonably well:
            - Tucker Lewis index (TLI) = 0.924 (≥ 0.90 is acceptable)
            - Comparative Fit Index (CFI) = 0.942 (≥ 0.90 is acceptable)
            - Root Mean Square Error of Approximation (RMSEA) = 0.026 (< 0.05 indicates good fit)
    - We will include the results in our full submission

- **Issue 2:** Using correlations to support validity.
    - **Developer Response 2:** We thank the panel members for their comments related to correlations. We identified four related issues. To address these issues, we added information below to (1) provide theoretical/hypothetical context for the reported Pearson correlation coefficients, (2) correlate measures with external data, (3) report complete correlation results with proper corrections and (4) provide information about #PCP-1 (Community Job Goal), #PCP-3 (ADL Goal), #CI-1 (Social Connectedness), and #CI-3 (Transportation Availability Scale). We articulated directional hypotheses for expected associations among measures and only tested those hypotheses. State-level socioeconomic measures derived from 2018 data provided by the Census Bureau were included in the hypothesis testing. All 14 measures were supported in at least one hypothesis.

      The table below lists our hypotheses and their test results. All hypotheses were directional; one-tailed tests were conducted.

Measure(s)	Relational Hypothesis	Test Results
Community Job Goal (PCP-1)	Urban settings provide a broader range of employment opportunities and hence, a larger choice of the types of jobs that are available for people with IDD. Urbanized states would be expected to find it easier to meet individuals' need for employment by including their wish for employment in their service plans.	Percentage of a state's population living in urban areas is positively and significantly correlated with PCP-1. r = 0.395 p = 0.011
Community Job Goal (PCP-1)	State Employment Leadership Network (SELN) was established in 2006 to support state public managers to offer expanded community-based employment options for people with IDD. We would expect SELN member states to have greater capacity and incentives to support the employment needs of their service participants and hence, to score higher on this measure.	Mean PCP-1 score for SELN member states is 0.4146 compared to a mean score of 0.3347 for non-member states. A one-tailed t-test of the difference between means yields a p-value of 0.055. Keeping in mind the low sample size (37 states), this result provides some support for the hypothesis.
Person-Centered Goals (PCP-2) and Chose Staff (CC-1)	State IDD service systems where it is common practice to develop a service plan based on the individual's preferences is likely to also provide staffing options based on their preferences.	These two measures are positively and significantly correlated. r = 0.344 p = 0.023

Measure(s)	Relational Hypothesis	Test Results
ADL Goal (PCP-3) and Community Job Goal (PCP-1)	In the support delivery system, Activities of Daily Living is a foundational element of assessment of functional support need, which is used to establish eligibility for service. Deficits in the ability to perform ADLs are therefore considered "low hanging fruit" in terms of developing a support plan. In many cases, ADL Goals may be carried over year over year for an adult receiving services, and it is unclear if they have chosen that goal. We would expect that service systems that use a deficit-based assessment and service planning may be more likely to have ADL goals in service plans. Person-centered-plans that are more progressive seek to support adults with IDD in their community employment and community participation goals, regardless of ADL deficits. We would expect to see community employment goals associated with a person's desire for community employment in progressive state service systems. We hypothesize a negative correlation between PCP-3 (ADL goal) and PCP-1 (Community job goal)	These two measures are negatively and significantly correlated. r = -0.342 p = 0.024
Lifelong Learning (PCP-4) and Has Friends (CI-2)	People with wider social circles are more likely to get exposed to new ideas and concepts. Therefore, states where people with IDD are more likely to report having friends (outside family and staff) would be expected to also have a high proportion of people with IDD reporting opportunities for lifelong learning.	These two measures are positively and significantly correlated. r = 0.764 p < 0.001
Satisfaction with Community Inclusion Scale (PCP-5) and Lifelong Learning (PCP-4)	Exposure to new ideas and concepts would be expected to increase one's expectations of inclusion in a broader range of community activities, thus increasing the sense of "relative deprivation" and increasing dissatisfaction with one's current level of community inclusion. We would therefore expect a negative association between PCP-5 and PCP-4	These two measures are negatively and significantly correlated. r = -0.498 p = 0.002

Measure(s)	Relational Hypothesis	Test Results
Social Connectedness (CI-1) and Respect for Personal Space (HLR-1)	People whose personal space is respected by the people with whom they interact are more likely to feel socially connected to them. We would therefore expect a positive association between social connectedness and respect for personal space	These two measures are positively and significantly correlated. r = 0.387 p = 0.012
Transportation Availability Scale (CI-3) and Satisfaction with Community Inclusion Scale (PCP-5)	People who have readily available means of transportation are more likely to be satisfied with their level of engagement in activities outside the home. CI-3 and PCP-5 would therefore be expected to be positively associated.	These two measures are positively and significantly correlated. r = 0.404 p = 0.009
Community Inclusion Scale (CI-4)	Resource-rich states would be expected to have greater ability to support people with IDD to engage in activities outside their home. Two measures of state-level resource availability were used to test this hypothesis: per-capita income and per-capita number of jobs.	Both measures of state-level resources are positively and significantly correlated with CI-4. Per-capita income: r = 0.345 p = 0.023 Per-capita number of jobs: r = 0.471 p = 0.003
Can Change Case Manager (CC-2) and Life Decisions Scale (CC-4)	A state system that allows its clients a high degree of choice in life decisions is expected to also allow them to choose or to change their case managers, given that these two areas of choice reflect a common service philosophy.	These two measures are positively and significantly correlated. r = 0.349 p = 0.022
Can Stay Home When Others Leave (CC-3) and Life Decisions Scale (CC-4)	A state system that allows its clients a high degree of choice in life decisions is expected to also provide them with the option of staying home alone when others leave, given that these two areas of choice reflect a common service philosophy.	These two measures are positively and significantly correlated. r = 0.552 p = 0.001

In conclusion, all 14 measures have an association with at least one other measure in line with theoretical expectations. In addition to expected associations with each other, hypothesized associations with measures based on external data are also supported. These findings provide evidence of validity at the measured entity level.

- **Issue 3:** There were concerns regarding whether or not the submitted measures should be considered PRO-PMs.

  Developer Response 3: We thank the panel members for the opportunity to discuss this important point. One panel member noted "… these are not outcome measures… these are things that are done or not done – processes of care", "Responses to some of the measures can be easily influenced by environmental factors … that are external to the support programs being measured". To address these concerns, we provide this further information for the panel members' re-consideration of the proposed measures as Person-Reported Outcome Performance Measures (PRO-PMs) in the context of Home and Community-Based Services (HCBS).

  In considering whether our proposed measures are PRO-PMs, we consulted the following definitions, featured in NQF publication <u>*Patient Reported Outcomes (PROs) in Performance Measurement*</u> (NQF, 2013):

  "Patient-reported outcomes (PROs) are defined as 'any report of the status of a patient's (or person's) health condition, health behavior, or experience with healthcare that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else.'" (p.5)

  "A PRO-based performance measure (PRO-PM) is based on PRO data aggregated for an entity deemed as accountable for the quality of care or services delivered. Such entities can include (but would not be limited to) long-term support services providers, hospitals, physician practices, or accountable care organizations (ACOs). NQF endorses PRO-PMs for purposes of performance improvement and accountability…" (p.5)

  In fact, in Table 1 (p.5), our project (National Core Indicators) was used as an example for PRO-PM, which affirms NQF's view on HCBS outcomes as PRO-PMs.

  The National Quality Forum (NQF) defines HCBS as "an array of services and supports delivered in the home or other integrated community setting that promote the independence, health and well-being, self-determination, and community inclusion of a person of any age who has significant, long-term physical, cognitive, sensory, and/or behavioral health needs" (NQF 2016). Measures of the quality of these services must therefore include outcomes within a broad range of life domains.

  In the context of quality monitoring, the proposed measures are using person-reported data to assess the extent to which people who are in receipt of funded services are experiencing quality life outcomes. In a discussion of person-reported outcomes in HCBS, Lipson (2019) points out that advances in the field of disability have broadened the understanding of quality of life and how it is measured from an individual's perspective. Quality domains central to quality of life include choice and satisfaction with residential settings, as well as addressing barriers to community participation such as limitations in transportation.

  To conclude, we contend that the submitted measures qualify as PRO-PMs in the context of HCBS. We acknowledge the panel member's comment that "responses to some of the measures can be easily influenced by environmental factors … that are external to the support programs being measured". For the measures we have put forward, however, we suggest that effective and flexible home and community-based supports can be developed to address environmental factors that may be serving as barriers. Measures that address individual choice and key life outcomes such as employment and community access reflect on quality of HCBS and are reported at the person-level as personal outcomes. Performance measures that give credit to HCBS providers for successfully overcoming environmental barriers to independent living and community integration should be given consideration for inclusion in measurement systems.

- **Issue 4:** Panel members sought additional evidence of validity through connection to other measures of quality.
    - **Developer Response 4:** We thank the panel members for their comments and acknowledge that there is a lack of presentation of external evidence in our original MIF and testing form to support the validity of the submitted measures and instruments. Here we include additional information for consideration.
    - Measures from National Core Indicators have been cross-walked and tested for their applicability for benchmarking, assessing, quality monitoring and comparing progress at various levels and contexts. Below list some of the external evidence by state, national and international level.
    - **At the state level**

State	Evidence
Arizona	Increased provider rates to incentivize Community and Supported Employment initiatives; Created District Employment Specialist positions, showing that components of #PCP-1 is relevant in the state's policymaking. (Bradley, Hiersteiner, &Bonardi 2016)
Massachusetts	The state department of developmental services Licensure and Certification data, an external state-level data source, clearly corroborates with NCI measure #CI-1 Social Connectedness. This was referenced by the state in a brief analysis report: Quality is No Accident. https://shriver.umassmed.edu/wp-content/uploads/2020/07/QINA-Friendship_final_web2.pdf
Kentucky	In 2010, the Kentucky Division of Developmental and Intellectual Disabilities implemented changes related to NCI measure #PCP-1 Community Job Goal and #CI-1 Social Connectedness, showing the relevance of those measures.
Many states, such as Tennessee and Michigan	Convene committees and quality improvement councils to review NCI data, which includes the submitted NCI measures.

- **At the National Level:** The National Core Indicators Measures selected for submission have demonstrated validity through alignment with multiple quality monitoring frameworks and tools, detailed below:
    - Medicaid Adult Core Health Care Quality Measure Set: In 2019, the Center for Medicaid and CHIP Services (CMCS) announced updates to the Medicaid Adult Core Health Care Quality Measure Set to include use of the National Core Indicators® (NCI®) to measure the quality of healthcare provided to adult Medicaid recipients on three measures [Life decisions scale (#CC-4); Transportation measure (component of #CI-3); everyday choices scale (includes #CC-1 and #CC-2)] that were selected to be reported to CMS this year. The Adult Core Set Measures are available at https://www.medicaid.gov/medicaid/quality-of-care/performance-measurement/adult-and-child-health-care-quality-measures/adult-health-care-quality-measures/index.html.
    - Recommended Measure set for Medicaid-Funded Home and Community Based Services: In 2020, CMS proposed a measure set for quality monitoring. Each of the measures in this submission were included as part of the CMS-proposed measure set. Detailed information is available at the following link. HCBS Recommended Measure Set RFI (medicaid.gov)
    - Medicaid Scorecard: NCI was one of the three experience-of-care surveys included in the Medicaid Scorecard for LTSS, which is used by CMS to increase public transparency and accountability about the state Medicaid programs' administration and outcomes. Details here: https://www.medicaid.gov/state-overviews/scorecard/state-use-patient-surveys-ltss-beneficiaries/index.html
    - HCBS Advocacy Coalition's inclusion of NCI in Settings Rule Monitoring: The Medicaid Home and Community Based Services (HCBS) Settings Rule, issued by CMS in 2014, requires states to engage in ongoing monitoring throughout implementation. The #CI-4 Community Inclusion Scale was recommended in a white paper as a monitoring tool for this purpose. Further details are available at https://hcbsadvocacy.org/2020-outcomes-paper/.
  - **At the International Level:** Measures from National Core Indicators have been cross-walked and tested for their applicability for benchmarking, assessing, and comparing progress towards a more inclusive society as described in the United Nations Convention on the Rights of People with Disabilities (UNCRPD), which includes a provision for the monitoring of outcomes of people with disability (Articles 31 and 33). To that end, National Core Indicators has been identified as providing a potential pathway to measurement of outcomes for people with IDD along key domains. Ticha et al (2018) laid out a conceptual framework for alignment of the UNCRPD with National Core Indicators. As a follow up study, Houseworth's analysis (Houseworth et al, 2019) was expanded to empirically test the framework and groupings of National Core Indicators by Articles of the UNCRPD. The results of our factor analysis largely aligned with Tichá et al.'s (2018) grouping of NCI items by UNCRPD article.
- In sum, all these works establish that the submitted measures have strong face validity, are widely recognized for relevance, and corroborated with externally sourced data.

- **Issue 5:** Exclusion criteria were unclear and inconsistent on MIF and testing forms. One panel member noted that "…the MIF notes a number of exclusions. However, the testing form checks the box for 'no exclusions'…"
    - **Developer Response 5:** We acknowledge the misalignment between testing form section 2b2.(exclusion analysis) and MIF sections about exclusions(s.8, s.9, s16), and would like to clarify and provide additional information.
    - Before we clarify about the exclusion criteria, some important context: To facilitate and accommodate person-centered reporting, the data collection instrument is divided into two sections, denoted by Roman numerals I and II. Section I of the survey contains questions about personal experiences and therefore may only be answered by the individual receiving developmental disabilities services. Section II of the survey---featuring questions about topics such as community involvement, choices, rights, and access to services—allows for responses from a "proxy," defined as a person who knows the individual well (such as a family member or friend).
    - At the end of Section I, the surveyor assesses whether the respondent appears to understand at least one question and answers in a cohesive manner. This assessment is the only subjective process in the exclusion determination process, but it is not done on an arbitrary or state-by-state basis. Rather, it is based on a protocol, included in the manual and reviewed during surveyor trainings, that apply uniformly to all surveyors across different participating states. The protocol is straightforward—the section must be marked "valid" if at least one question in the section was answered in a manner that the basic level of comprehension was shown, and a clear response given either verbally (e.g., yes/no) or non-verbally (nodding/shaking head).
    - NCI and participating states routinely conduct surveyor training and surveyor shadowing and reviewing processes that ensure, among other things, that surveyors are applying this assessment (whether or not Section I was valid) strictly based on the protocol.
    - A proxy is not required, and sometimes no proxy is available, so the person with disabilities may answer both Section I and Section II (which is important for criteria c below).
    - There are 4 section-based exclusion criteria:
      For Section I items:
      (a) Based on survey protocol, the surveyor found that the respondent did not give any valid responses to any Section I questions, or
      (b) All questions in Section I were left blank or marked "not applicable" or "don't know".
      For Section II items:

      (c) Section II was completed without using a proxy, while Section I was deemed invalid (see criteria a above), or

      (d) All questions in Section II were left blank or marked "not applicable" or "don't know".
      Here is the distribution of exclusions among states:

Exclusion criterion	N excluded	% excluded	Distribution across states N=22,009 (Min, 25th, 50th, 75th percentile, and max)
For Section I items: (a) Based on survey protocol, the surveyor found that the respondent did not	5,053	22.9%	(3%, 13%, 23%, 26%, 62%)

Exclusion criterion	N excluded	% excluded	Distribution across states N=22,009 (Min, 25th, 50th, 75th percentile, and max)
give any valid responses to any Section I questions			
(b) All questions in Section I were left blank or marked "not applicable" or "don't know".	1,882	8.6%	(0%, 0%, 5%, 14%, 42%)
For Section II items: (c) Section II was completed without using proxy, while Section I was deemed invalid (see criteria a above), or	59	0.3%	(0%, 0%, 0%, 0%, 5%)
(d) All questions in Section II were left blank or marked "not applicable" or "don't know". Here is the distribution of exclusions among states	311	1.4%	(0%, 0%, 1%, 1%, 7%)

Interpretations:

- Exclusion (a), (the surveyor found that the respondent did not give any valid responses to any Section I questions), accounting for 22.9% of all surveys, represents the majority (69%) of all exclusions and is meant to safeguard the validity of measures that utilize Section I items. This exclusion, with its conservative approach, prevents the inclusion of responses with sub-standard reliability in measure calculations.
- Exclusion (b) (All questions in Section I were left blank or marked "not applicable" or "don't know") accounts for about 8.6% of all surveys and represents about a quarter (26%) of all exclusions. It is purely objective and is needed to prevent the inclusion of responses that do not contribute meaningful data for Section I items.
- Exclusion (c) (Section II was completed without using proxy, while Section I was deemed invalid) only accounts for 0.3% of all surveys and represents less than 1% of all exclusions. However, it is in place to safeguard the validity and reliability of measures that utilize Section II items by excluding responses provided by an individual whose responses to Section I were assessed as unreliable. Given that a very small percentage of individuals are being excluded, it is unlikely this exclusion unduly affects the measure score.
- Exclusion (d) (All questions in Section II were left blank or marked "not applicable" or "don't know") accounts for 1.4% of all surveys, 4% of all exclusions. Its determination is purely objective and is needed to prevent the inclusion of surveys that do not contribute meaningful data for Section I items.
- One panel member noted that answers of "unknown" or "not applicable" are dropped from the denominator. We thank the panel member for noting the lack of clarity on this and will amend our numerator and denominator statements in the full submission. Answers of "unknown" or "not applicable" do not get included because including such answers in the calculations would cause underestimation—for example, those who already have a job would

not have a job-related goal in their HCBS service plan and would answer "not applicable" to "would you like to have a job in the community". By including such answers, the #PCP-1 Community Job Goal calculations would have lower rates than otherwise, thereby masking the true gap (those who want a job but do not have a job goal) in quality monitoring.
    - In conclusion, exclusions are based on the uniformly applied criteria, most of which are objective and all of which are standardized. The exclusions were put in place to ensure accurate calculation of the measures and to safeguard validity and reliability. It is important to note that to the extent possible, exclusions eliminate unreliable responses, not the entire survey. For example, a survey where Section I responses are excluded from measure calculations may still be included in measures based on Section II items if Section II responses were provided by a proxy. We intend to amend the testing attachment section 2b2 to align with MIF s.8, s.9 and s.16.
- **Issue 6:** States cherry-pick favorable surveys or survey sites.
- **Developer Response 6:** States' survey strategies are determined by workplans, which are third-party designed and reviewed. Many states contract with surveying agencies to conduct surveys. States do not get to pick "successful" sites or programs for interviewing. The National Association of State Directors of Developmental Disabilities Services (NASDDDS) provides general oversight and guidance in all states NCI activities. All state HCBS eligible populations are generally included in the survey frame unless reasonable justifications can be made.
    - References:
    - Bradley, V.J., Hiersteiner, D., Bonardi, A. (2016). A focus on System-Level Outcome Indicators. In Cross Cultural Quality of Life: Enhancing the Lives of People with Disability. Schalock and Keith Eds. American Association on Intellectual and Developmental Disability.
    - Houseworth, J., Stancliffe, R., & Tichá, R. (2019). Examining the National Core Indicators' Potential to Monitor Rights of People with Intellectual and Developmental Disabilities According to the CRPD. Journal of Policy and Practice in Intellectual Disabilities, 16(4), 342-351
    - Lipson D.J. (2019), Person-Reported Outcome Measures for Home and Community-Based Services. HCBS Quality Measures Issue Brief. Mathematica https://www.medicaid.gov/medicaid/quality-of-care/downloads/hcbs-quality-measures-brief-2-person-reported-outcome.pdf
    - National Quality Forum (2013). Patient Reported Outcomes (PROs) in Performance Measurement. Final Report. Washington, DC: NQF, January 2013. Available at https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=72537
    - National Quality Forum (2016). Quality in Home- and Community-Based Services to Support Community Living: Addressing Gaps in Performance Measurement. Final Report. Washington, DC: NQF, September 2016. Available at https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=83433
    - Tichá, R., Qian, X., Stancliffe, R.J., Larson, S., & Bonardi, A. (2018). Alignment between the Convention on the Rights of Persons with Disabilities and the National Core Indicators. Journal of Policy and Practice in Intellectual Disabilities, 15(3), 247-255.

Developer Submission

NQF #: 3622

Corresponding Measures:

De.2. Measure Title: National Core Indicators for Intellectual and Developmental Disabilities (ID/DD) Home- and Community-Based Services (HCBS) Measures

Co.1.1. Measure Steward: Human Services Research Institute

De.3. Brief Description of Measure: National Core Indicators for Intellectual and Developmental Disabilities Home- and Community-Based Services Measures ("NCI for ID/DD HCBS Measures" hereafter) originate from NCI(R) In-Person Survey (IPS), an annual multi-state cross-sectional survey of adult recipients of state developmental disabilities systems´ supports and services. First developed in 1997 by the National Association of State Directors of Developmental Disabilities Services (NASDDDS) in collaboration with Human Services Research Institute (HSRI), the main aims of NCI for ID/DD HCBS Measures were to evaluate person-reported outcomes and assess state developmental disabilities service systems performance in various domains and sub-domains accordingly. The unit of analysis is "the state", and the accountable entity is the state-level entity responsible for providing and managing developmental disabilities services. Currently, 46 states and the District of Columbia are members of the NCI program. To align with member states´ fiscal schedules, the annual survey cycle typically starts on July 1 and ends on Jun 30 of the following year.

Gathering subjective information and data from people with ID/DD poses unique challenges due to potential intellectual and developmental limitations experienced by the population. As such, extensive work went into the processes of developing NCI IPS administration methods, survey methodology and measure design and revisions. The original development built on direct consultation with members of the target population and their advocates, as well as extensive literature review and testing.

The NCI for ID/DD HCBS Measures consist of 14 measures in total, including:

Five measures in the HCBS Domain: Person-Centered Planning (PCP) and Coordination

#PCP-1 The proportion of people who express they want a job who have a related goal in their service plan (Community Job Goal)

#PCP-2 The proportion of people who report their service plan includes things that are important to them (Person-Centered Goals)

#PCP-3 The proportion of people who express they want to increase independence in functional skills (ADLs) who have a related goal in their service plan (ADL Goal)

#PCP-4 The proportion of people who report they are supported to learn new things (Lifelong Learning)

#PCP-5 The proportion of people who report satisfaction with the level of participation in community inclusion activities (Satisfaction with Community Inclusion Scale)

Four measures in the HCBS Domain: Community Inclusion

#CI-1 The proportion of people who reported that they do not feel lonely often (Social Connectedness)

#CI-2 The proportion of people who reported that they have friends who are not staff or family members (Has Friends)

#CI-3 The proportion of people who report adequate transportation (Transportation Availability Scale)

#CI-4 The proportion of people who engage in activities outside the home (Community Inclusion Scale)

Four measures in the HCBS Domain: Choice and Control

#CC-1 The proportion of people who reported they chose or were aware they could request to change their staff (Chose Staff)

#CC-2 The proportion of people who reported they could change their case manager/service coordinator (Can Change Case Manager)

#CC-3 The proportion of people who live with others who report they can stay home if they choose when others in their house/home go somewhere (Can Stay Home When Others Leave)

#CC-4 The proportion of people who report making choices (independently or with help) in life decisions (Life Decisions Scale)

And one measure in the HCBS Domain: Human and Legal Rights

#HLR-1 The proportion of people who report that their personal space is respected in the home (Respect for Personal Space Scale)

1b.1. Developer Rationale: In the past 30 years, we have witnessed significant changes in the nature and extent of the publicly financed systems of services and supports for individuals with developmental disabilities and their families. Growing service needs, declining state revenues, mounting pressures on federal and state Medicaid budgets, and heightened federal quality management expectations are placing increasing demands on state developmental disabilities agencies to improve accountability, more effectively track outcomes, and strengthen their capacity to assess and improve service quality and responsiveness. The initiation of the National Core Indicators collaboration 20 years ago was in direct response to these and other forces described below. Public systems that provide home and community based services for people with intellectual and developmental disabilities are highly diverse and widely dispersed. As the places where people live and work and the supports that they receive become more individualized, the necessity for effective performance appraisal and quality improvement systems which directly capture people's experience of services and outcomes has continued to grow. The administration of the Home and Community Based Medicaid Waiver program by the U.S. Department of Health and Human Services Centers for Medicare and Medicaid Services (CMS) requires to states to monitor and provide assurances of the quality and performance of their support systems. Federal expectations included in the revised formal Waiver application require that each state agency explain the operational features of its planned HCBS Medicaid Waiver program in detail. States also describe the components of the comprehensive quality management strategy that they intend to employ to monitor and improve the accessibility and quality of services offered to Waiver participants and to demonstrate that required Waiver assurances are being met. Through this approach, CMS has information allowing assessment of a state´s quality management strategy based on the presence or absence of objective, measurable data demonstrating that the state is actively managing the quality of Medicaid Waiver services and supports through a continuous process of discovery, remediation, and improvement. To respond to the emphasis on performance and outcome data, CMS has recommended that states consider the use of standardized tools, such as the National Core Indicators, to gather and analyze information on Waiver participants´ outcomes and satisfaction with the services they receive. The changing expectations on the part of CMS regarding quality monitoring and performance measurement parallel gubernatorial and legislative initiatives in several states aimed at increasing government accountability, improving service outcomes, and strengthening program responsiveness

As PRO-PMs, these measures are designed to assist states in moving beyond compliance to assess service quality, in line with the minimum standards of the federal regulations. The full measure set is designed for states to measure their performance in the HCBS Domains address by the NQF 2016 report and establish goals and benchmarks for performance and leading states to focus on quality improvement, a step beyond compliance.

The 14 measures being submitted collectively addresses 4 of the measurement domains identified by the national, multi-stakeholder Committee on HCBS convened by NQF as areas where "performance measures are needed to drive systems change, tie performance to outcomes, allow consumers to make informed choices, and compare the effectiveness of different models of HCBS and of HCBS versus institutional services." (NQF, 2016, p. 2).

HCBS Domain: Person-Centered Planning (PCP) and Coordination (Measures #PCP-1 – #PCP-5)

CMS final rules on HCBS Waiver Requirements, published in 2014, promulgated expectations that waiver states establish a Person Centered Planning Process and the Person Centered Plan within a transition period now extended until 2023. The 5 measures in the PCP Domain submitted for NQF endorsement have been developed in response to state DD agencies' requests for measures which can demonstrate the required compliance with both the person centered planning process, and the HCBS Settings requirements. During annual meetings of the NASDDDS members (State Directors of Developmental Disabilities Services), requests were made to develop and /or adapt existing measures to assist states in determining their performance related to these new rules.

HCBS Domain: Community Inclusion (Measures #CI-1 – #CI-4)

The Americans with Disabilities Act (ADA) of 1990, and subsequent legal actions such as the 1999 Supreme Court decision in the landmark Olmstead case, have established full integration with society as a basic human right for individuals with disabilities. Thus, the adequacy and quality of state DD services depends on the extent to which they support their consumers' community inclusion. The 4 measures in the Community Inclusion Domain submitted for NQF endorsement measure the success with which state service systems support individuals with IDD in participating in the activities outside their home and integrating with their communities. Given that community inclusion was identified by the NQF Committee on HCBS as one of the key outcomes expected of a service system, these measures are also vital in comparing competing HCBS models.

HCBS Domain: Choice and Control (Measures #CC-1 – #CC-4)

The main purpose of HCBS is to promote and maintain the wellbeing and quality of life of people with functional limitations. Ability to make life choices such as where and with whom to live, services they receive, and how those services are delivered, have been shown to improve the quality of life among people with disabilities (Willis, Grace, & Roy, 2008; Stancliffe et al., 2011). Therefore, measures that assess the level of choice and control that individuals with IDD have over their life and their services are crucial performance measures for state service systems. Monitoring changes in their measures of choice and control will help state service systems achieve continuous quality improvement. The measures are also vital in comparing and improving HCBS systems.

Stancliffe, R.J., Lakin, K.C., Larson, S., Engler, J., Taub, S., & Fortune, J. (2011). Choice of living arrangements. Journal of Intellectual Disability Research, 55(8), 746-762.

Willis, D., Grace, P. J., & Roy, C. (2008). A central unifying focus for the discipline: Facilitating humanization, meaning, choice, quality of life, and healing in living and dying. Advances in Nursing Science, 31(1), 28-40.

HCBS Domain: Human and Legal Rights (Measure #HLR-1)

Within this domain, we are submitting a measure of the respect for one's privacy. This is an important component of basic human rights and dignity that all members of society are entitled to. The Convention on the Rights of Persons with Disabilities promulgates the protection of basic human rights for people with disabilities. Therefore, the extent to which individuals with IDD report that their personal space is respected is an indicator of the ability of an HCBS system to protect its consumers' basic human rights and dignity. It is, therefore, a crucial measure of service quality. Monitoring changes in their scores on this measure will enable state service systems to make the necessary changes to their services to avoid intrusiveness and instead facilitate privacy and dignity among individuals with disabilities.

S.4. Numerator Statement: The NCI for ID/DD HCBS Measures use values between 0 and 1 as the scores. Typically, the numerator is the number of respondents who selected the most positive response category (e.g. "yes", "always"). The attached file SuppTable_Measures_210420_508.xlsx lists what constituted the most positive response categories for each measure item, as well as other detailed information as relevant for S.2b.

S.6. Denominator Statement: For each measure, the denominator is the number of respondents (adult recipients of state developmental disabilities services) who provided valid answers to the respective survey question, except those that meet the exclusion criteria (see S.8. below for details).

If the denominator for a state is fewer than 20, the measure score is censored to protect the confidentiality of respondents.

S.8. Denominator Exclusions: At the end of Section I, the surveyor assesses whether the respondent appears to understand at least one question and answers in a cohesive manner. This assessment is the only subjective process in the exclusion determination process, but it is not done on an arbitrary or state-by-state basis. Rather, it is based on a protocol, included in the survey manual and reviewed during surveyor trainings, that apply uniformly to all surveyors across different participating states. The protocol is straightforward—the section must be marked “valid” if at least one question in the section was answered in a manner that the basic level of comprehension was shown, and a clear response given either verbally (e.g. yes/no) or non-verbally (nodding/shaking head). NCI and participating states routinely conduct surveyor training and surveyor shadowing and reviewing processes that ensure, among other things, that surveyors are applying this assessment (whether or not Section I was valid) strictly based on the protocol. If the surveyor´s assessment is that Section I is not valid, the respondent´s Section I data are flagged for exclusion from the numerators and denominators. However, the individual is not removed from the dataset.

- If Section I data are excluded, Section II data are flagged for exclusion from the numerators and denominators -unless- a proxy respondent was used in Section II. If the respondent or proxy did not answer any questions in Section II, the survey is removed from the denominators of Section II items.

Responses are excluded from numerators and denominators for Section I items if:

(a) The surveyor indicated that the respondent did not give consistent and valid responses, or

(b) All questions in Section I were left blank, or marked "not applicable" or "don’t know".

Responses are excluded from numerators and denominators for Section II items if:

(a) the individual receiving supports was marked as the sole respondent to all questions in Section II but Section I was deemed invalid, or

(b) All questions in Section II were left blank, or marked "not applicable" or "don’t know".

For each measure item, missing responses and responses indicating "not applicable" or "don’t know" were excluded from denominators. The distribution of exclusions among states is shown in Testing Attachment 2b2.2. Please see S.9. for more details on denominator exclusions.

De.1. Measure Type: Outcome: PRO-PM

S.17. Data Source: Instrument-Based Data

S.20. Level of Analysis: Population : Regional and State

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?

1. Evidence and Performance Gap – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

NQF_evidence_attachment_NCI_210420_508-637545337274869400.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?
Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:**

Measure Title: National Core Indicators for Intellectual and Developmental Disabilities (ID/DD) Home- and Community-Based Services (HCBS) Measures

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:

Date of Submission: **4/9/2021**

1a.1. This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☒ Outcome:

  ☒Patient-reported outcome (PRO): **Experience with Care**

  PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)

☐ Intermediate clinical outcome (*e.g., lab value*):

☐ Process:

  ☐ Appropriate use measure:

☐ Structure:

☐ Composite:

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

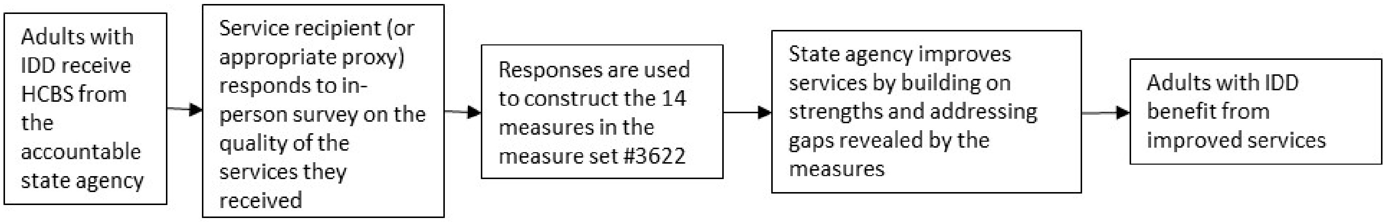
The development and testing of the measures were shaped by the conceptual framework and theory of quality measurement set forth in NQF's 2016 report on Home and Community-Based Services (HCBS).[1] Instead of being derived from systems or clinical data, the submitted measures directly reflect the often-marginalized and ignored voices of those who receive publicly-funded HCBS. Originally developed in 1997 and currently being used in 46 states around the country and Washington D.C., the submitted measures help the state agency responsible for serving people with intellectual/developmental disabilities (IDD) identify service

---

[1] National Quality Forum. (2016). Quality in Home and Community-Based Services to Support Community Living: Addressing Gaps in Performance Measurement. Retrieved from https://www.qualityforum.org/Publications/2016/09/Quality_in_Home_and_Community-Based_Services_to_Support_Community_Living__Addressing_Gaps_in_Performance_Measurement.aspx

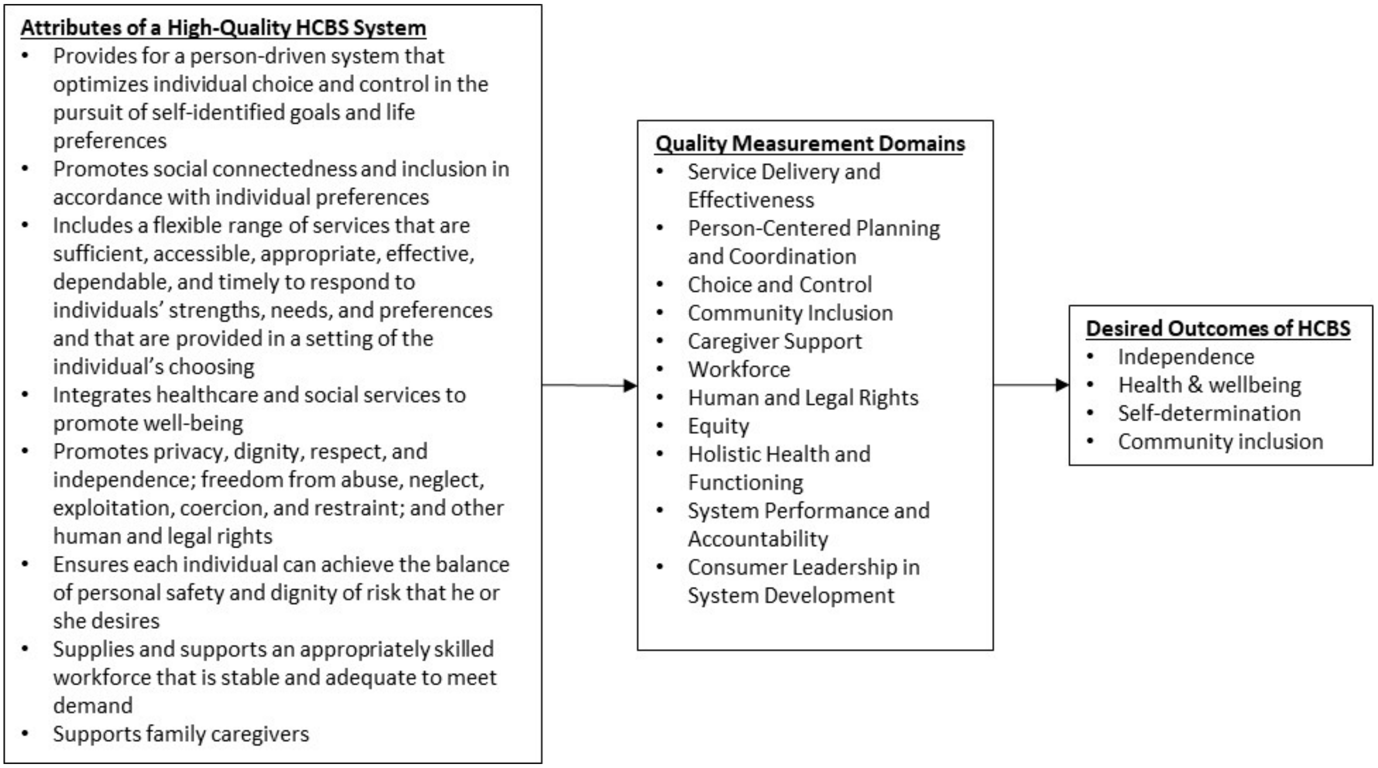
strengths and gaps, providing guidelines for improving the state's service systems. Figure 1 below describes the logic flow from service provision to service improvement through the use of the 14 measures.

Figure 1. Logic flow from HCBS provision to service quality improvement



All 14 measures fit into the measurement domains identified in the NQF HCBS Quality Framework[1] (also see the "Measure Descriptions" tab of the Supplemental Tables). The HCBS logic model formulated in the NQF report is summarized in the following diagram (Figure 2).

Figure 2. The NQF HCBS Framework, Domains, and Outcomes (NQF, 2016)



The logic diagrams that follow position each proposed measure within this overarching logic model. The leftmost box in each diagram represents an attribute of a high-quality HCBS system and can be interpreted as process indicators. The boxes to the right of the "process" box depict outcomes. In diagrams where there are two outcome boxes, the one in the middle depicts a more proximate outcome and the rightmost one depicts a more distal outcome. The outcomes depicted in the diagrams are those identified as desirable outcomes in the above overarching logic model.

Figure 3. Logic Model for PCP-1: Community Job Goal, PCP-2: Person-Centered Goals, and PCP-3: ADL Goal

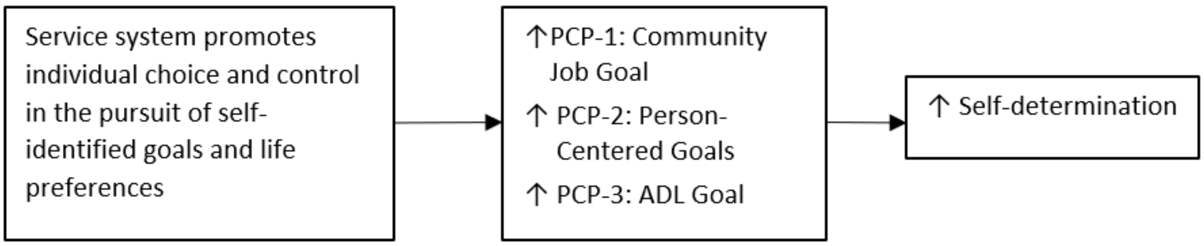
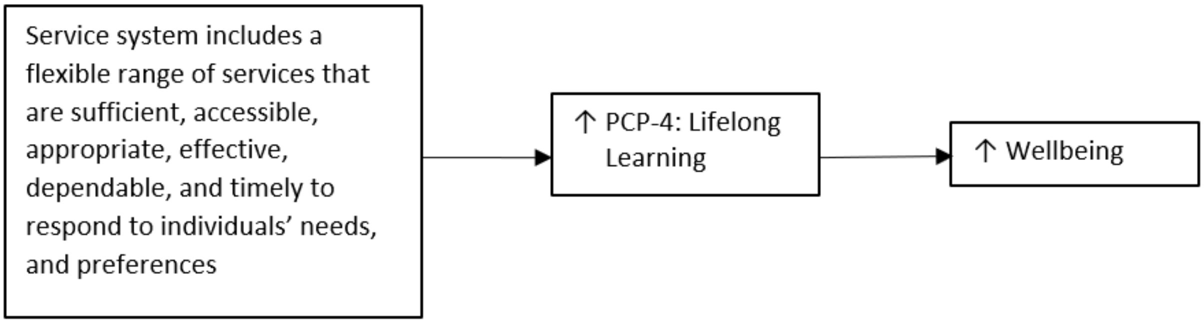


Figure 4. Logic Model for PCP-4: Lifelong Learning



Note: For evidence supporting the notion that lifelong learning promotes wellbeing for individuals with disabilities, see Merriam & Kee (2014)[2]

Figure 5. Logic Model for PCP-5: Satisfaction with Community Inclusion Scale

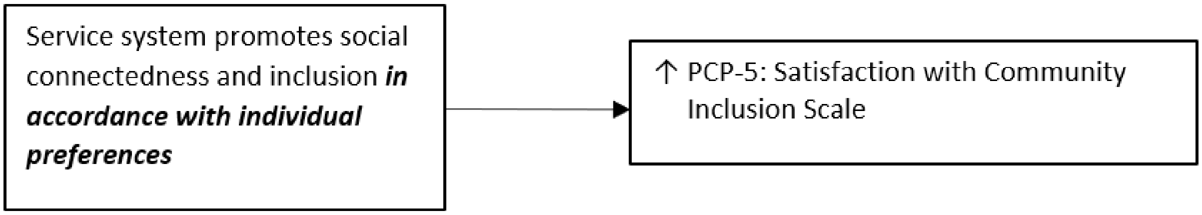


Figure 7. Logic Model for CI-1: Social Connectedness, CI-2: Has Friends, and CI-4: Community Inclusion Scale

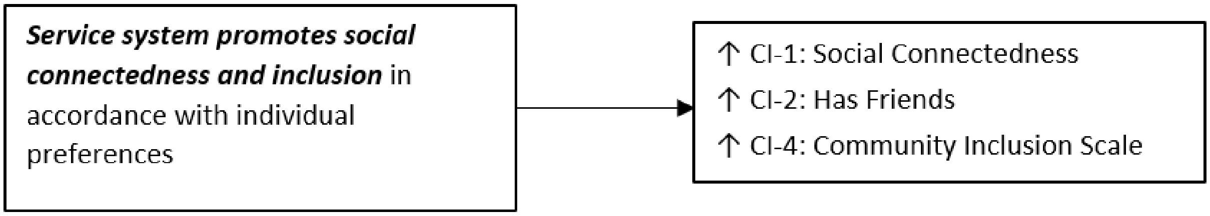
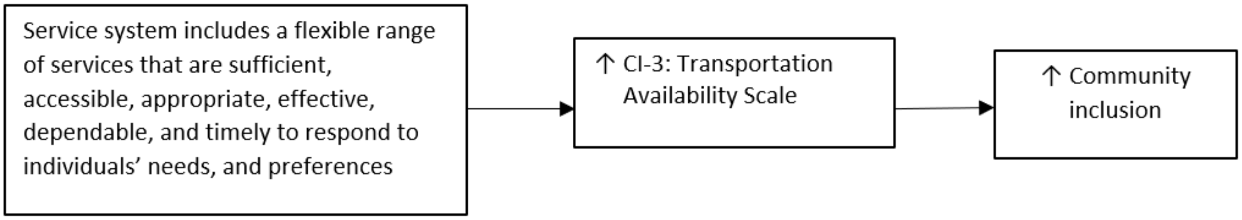


Figure 8. Logic Model for CI-3: Transportation Availability Scale



---

[2] Merriam, S.B. & Kee, Y. (2014). Promoting community wellbeing: The case for lifelong learning for older adults. *Adult Education Quarterly, 64*(2), 128-144. doi:10.1177/0741713613513633

Figure 9. Logic Model for CC-1: Chose Staff and CC-2: Can Change Case

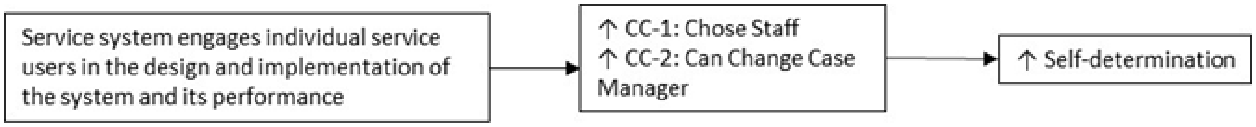


Figure 10. Logic Model for CC-3: Can Stay Home When Others Leave

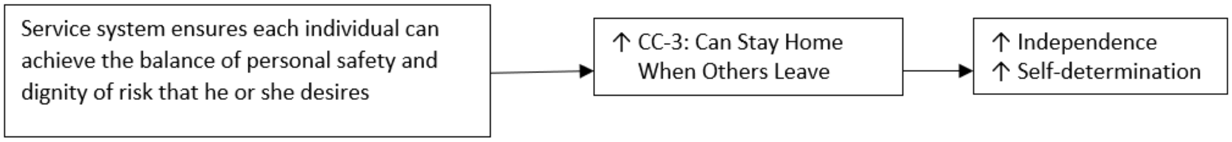


Figure 11. Logic Model for CC-4: Life Decisions Scale

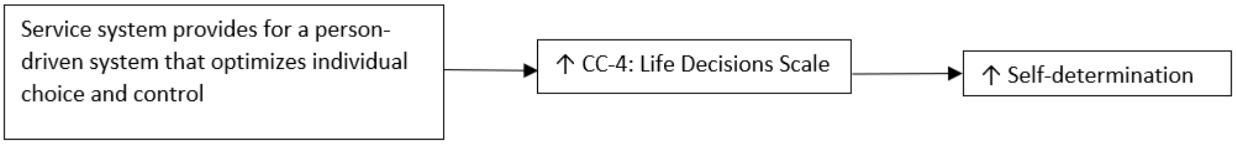
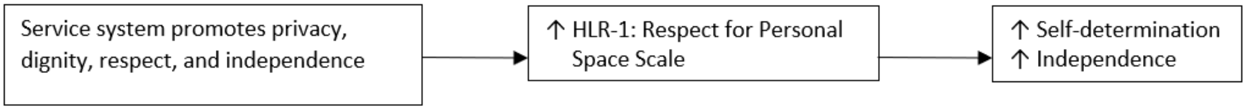


Figure 12. HLR-1: Respect for Personal Space Scale



1a.3 Value and Meaningfulness: **IF** this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

The measures submitted for endorsement are reflective of priorities that have been identified by the population with intellectual and developmental disabilities.

These priorities were brought forward through expert panelists who came together to establish the NQF framework for HCBS quality described earlier in this submission[3].

Researchers at University of Minnesota's Rehabilitation Research and Training Center on Outcome Measures conducted a study to establish the content and social validity of the NQF HCBS Quality Framework with stakeholders. As described in their research center's brief: "Involving Stakeholders to address challenges in HCBS Measure Development"[4], this was accomplished through a Participatory Planning and Decision-Making (PPDM) process. The PPDM process included meeting with all stakeholder groups and providing them with an opportunity to evaluate the NQF framework, add to it, and stipulate which personal outcomes and service characteristics were most important to measure. In order to obtain a nationally representative sample, PPDM groups were conducted across the country with each stakeholder group which included people with intellectual and developmental disabilities, mental health conditions, traumatic brain injury, physical

---

[3] National Quality Forum. (2016). Quality in Home and Community-Based Services to Support Community Living: Addressing Gaps in Performance Measurement. Retrieved from https://www.qualityforum.org/Publications/2016/09/Quality_in_Home_and_Community-Based_Services_to_Support_Community_Living__Addressing_Gaps_in_Performance_Measurement.aspx

[4] Rehabilitation Research and Training Center on Outcome Measurement. (RTC/OM). (2020) Brief 1: Involving Stakeholders to Address Challenges in HCBS Measure Development. Retrieved from: https://publications.ici.umn.edu/rtcom/briefs/brief-one-involving-stakeholders-to-address-challenges-in-hcbs-mesure-development

disabilities, and a variety of age-related conditions[5]. Additional focus groups were organized for family members, HCBS support providers, and groups of public managers.

Overall, results from PPDM groups conducted by the University of Minnesota Researchers indicated a high degree of stakeholder support for the content of the NQF HCBS Quality Measurement framework, further validating the framework to support quality improvement work. Stakeholders prioritized measures of Person-Centered Planning and Coordination, Choice and Control, and Human and Legal Rights.

Stakeholders in the study did provide input resulting in recommendations for number of revisions or additions to the NQF Framework for quality measurement. These included: (1) adding within the broad *community inclusion* domain a subdomain focused on access to and quality of ***transportation*;** (2) the addition of a stand-alone domain for ***employment***; and (3) a greater focus on ***the self-determination*** of people with disabilities rather than the degree of choice and control they experience.

The NCI for ID/DD HCBS Measures we are submitting are entirely responsive to priorities identified by the stakeholders through this study. Measures of choice and control which are submitted can be viewed as core elements that are supportive of self-determination, as detailed in the logic model section above.

---

[5] Rehabilitation Research and Training Center on Outcome Measurement. (RTC/OM). (2016). Stakeholder Input: Identifying Critical Domains and Subdomains of HCBS Outcomes. Retrieved from https://rtcom.umn.edu/phases/phase-1-stakeholder-input

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

- The monitoring of a range of measures withing quality domains in Home and Community Based Services through the National Core Indicators is an indication of the breadth of services for which states are held accountable for delivering quality services. For example, the state of California has passed legislation (effective 2019) requiring the public reporting of NCI measures to the governing boards of Regional developmental disabilities service administrative entities. Such reporting is intended to be used to identify needed improvements in services for people served by the system. Improvements are implemented at the state level or through developmental disabilities service administrative entities.
- California Welfare and Institutions Code 4571 specifies that regional centers present NCI data to their governing boards: https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=WIC&division=4.5.&title=&part=&chapter=4.&article=
- (b) The department, in consultation with stakeholders, shall identify a valid and reliable quality assurance instrument that assesses consumer and family satisfaction, provision of services in a linguistically and culturally competent manner, and personal outcomes. The instrument shall do all of the following:
- (1) Provide nationally validated, benchmarked, consistent, reliable, and measurable data for the department's Quality Management System.
- "h) (1) Each regional center shall annually present data collected from, and the findings of, the quality assurance instrument described in subdivision (b) for that regional center, at a public meeting of its governing board in order to assess the comparative performance of the regional center and identify needed improvements in services for consumers, including, but not limited to, case management services.

Additional outcomes, with specific state examples tied to specific domains of the HCBS quality framework are as follows.

Person-Centered Planning (PCP) and Coordination Domain:

#PCP-1 The proportion of people who express they want a job who have a related goal in their service plan (Community Job Goal)

#PCP-2 The proportion of people who report their service plan includes things that are important to them (Person-Centered Goals)

#PCP-3 The proportion of people who express they want to increase independence in functional skills (ADLs) who have a related goal in their service plan (ADL Goal)

#PCP-4 The proportion of people who report they are supported to learn new things (Lifelong Learning)

#PCP-5 The proportion of people who report satisfaction with the level of participation in community inclusion activities (Satisfaction with Community Inclusion Scale)

Service coordination training and monitoring to ensure that planning includes an exploration of work options and detailing that in the service plan.

- The MO Statewide Developmental Disabilities Agency demonstrated low rates people who express they want a job who have a related goal in their service plan (#PCP-1) in 2016. This led to an improvement initiative called **Employment First Collaborative**, and **Empowering Through Employment**. As a result of these initiatives, significant increases in the number of service plans with employment authorizations occurred between 2016 and

2020 (increase from 446 to 1023) See MO Quality Outcomes 2020 report, page 53, https://dmh.mo.gov/media/pdf/moqo-you-daily-living-employment-report

- Kentucky DD Services Quality Improvement Committee track performance on the measure of across 3 year cycles. The 2020 Committee Recommendations include specific interventions needed to accomplish improvement goals including enhanced information dissemination and service coordinator training. https://www.kentuckycq.org/wp-content/uploads/2020/12/QIC-recommendations-report-2020.pdf
- States review provider organizations for statutory assurances that service plan regulations are met, including reviews of person centered planning outcomes. For example, Indiana Bureau of Developmental Disability Services reports to CMS in their 1915(c) Appendix D Service Planning Waiver Performance Measures. Quality of service coordination, and the extent to which service plans reflect their goals are monitored. http://ai.org/fssa/files/CIH%20Final-Approved.pdf
- The Ohio Division of Developmental Disabilities website, includes an NCI Dashboard on Service Planning, "does your service plan include things that are important to you?" (#PCP-2) reported out by regions across the state for comparisons. https://dodd.ohio.gov/wps/portal/gov/dodd/about-us/dodd_data/nci/service_plan. Service coordinator capacity building in person-centered planning drives improvements in the inclusion of goals that are individualized and relevant to the person.
- States may track the extent to which personal preferences are reflected in community inclusion activities (#PC-5). The states of MO Developmental Disabilities agency monitors this measure Quality Outcomes #1 Daily Life: People Participate in Meaningful Daily Activities of their Choice. https://dmh.mo.gov/media/pdf/moqo-you-daily-living-employment-report

Community Inclusion Domain:

#CI-1 The proportion of people who reported that they do not feel lonely often (Social Connectedness)

#CI-2 The proportion of people who reported that they have friends who are not staff or family members (Has Friends)

#CI-3 The proportion of people who report adequate transportation (Transportation Availability Scale)

#CI-4 The proportion of people who engage in activities outside the home (Community Inclusion Scale)

- Massachusetts DDS quality assurance service intervention in 2016 reported on the reported proportion of people who do not feel lonely often (#CI-1) in its effort to improve supports to develop and maintain relationships with family and friends. The proportion of people who feel lonely was included in the report along with additional NCI indicators on friends and support to visit family and friends. https://shriver.umassmed.edu/wp-content/uploads/2020/07/DDS-Relationships_final-Jan-2016.pdf Massachusetts adopted a successful initiative in 2015 to increase friendships among people with I/DD with others who were not staff or family members. The initiatives were Identified as support strategies called Widening the Circle Mini-grants and "Building Friendships at work toolkits". https://archives.lib.state.ma.us/bitstream/handle/2452/783602/on1029840610.pdf?sequence=1&isAllowed=y https://shriver.umassmed.edu/wp-content/uploads/2020/07/DDS-Relationships_final-Jan-2016.pdf
- The Kentucky Developmental Disabilities Services Quality Improvement Committee uses the choice and control measure to track performance across 3-year cycles. The 2020 Committee Recommendations include interventions to accomplish the improvements needed to increase community inclusion including: service design changes, Medicaid Service definition changes to expand access to transportation, and changes to statewide workgroup engagement with service providers. https://www.kentuckycq.org/wp-content/uploads/2020/12/QIC-recommendations-report-2020.pdf
- 26 states included the measure of transportation adequacy (#CI-3) in their reporting through the Medicaid Adult Core Measure set in 2019. The portal for CMS 2019 Medicaid and CHIP Adult Core Measure set was not available at the time of this application but can be accessed through the CMS website after April 30, 2021.

Choice and Control Domain:

#CC-1 The proportion of people who reported they chose or were aware they could request to change their staff (Chose Staff)

#CC-2 The proportion of people who reported they could change their case manager/service coordinator (Can Change Case Manager)

#CC-3 The proportion of people who live with others who report they can stay home if they choose when others in their house/home go somewhere (Can Stay Home When Others Leave)

#CC-4 The proportion of people who report making choices (independently or with help) in life decisions (Life Decisions Scale)

- The District of Columbia used the indicator of staff choice (#CC-1) along with #CC-4 (Life decisions scale below) and other NCI data on Guardianship, to determine if the use of guardians impacted quality of life, specifically indicated by limited choice. The service intervention was directed at increasing the use of supported decision making in lieu of guardianship when appropriate. The Supported Decision-Making Project was launched in 2019 as a means to expand self-determination which includes greater choice and control for people who are receiving HCBS supports. https://dds.dc.gov/sites/default/files/dc/sites/dds/publication/attachments/NCI%20and%20Supported%20Decision%20Making_0.pdf

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Systematic Review	Evidence
Source of Systematic Review: • Title • Author • Date • Citation, including page number • URL	*
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	*
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	*
Provide all other grades and definitions from the evidence grading system	*
Grade assigned to the **recommendation** with definition of the grade	*
Provide all other grades and definitions from the recommendation grading system	*
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	*

Systematic Review	Evidence
Estimates of benefit and consistency across studies	*
What harms were identified?	*
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	*

*cell intentionally left blank

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 **Briefly SYNTHESIZE the evidence that supports the measure.** A list of references without a summary is not acceptable.

1a.4.2 **What process was used to identify the evidence?**

1a.4.3. **Provide the citation(s) for the evidence.**

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

If a COMPOSITE *(e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.*

In the past 30 years, we have witnessed significant changes in the nature and extent of the publicly financed systems of services and supports for individuals with developmental disabilities and their families. Growing service needs, declining state revenues, mounting pressures on federal and state Medicaid budgets, and heightened federal quality management expectations are placing increasing demands on state developmental disabilities agencies to improve accountability, more effectively track outcomes, and strengthen their capacity to assess and improve service quality and responsiveness. The initiation of the National Core Indicators collaboration 20 years ago was in direct response to these and other forces described below. Public systems that provide home and community based services for people with intellectual and developmental disabilities are highly diverse and widely dispersed. As the places where people live and work and the supports that they receive become more individualized, the necessity for effective performance appraisal and quality improvement systems which directly capture people’s experience of services and outcomes has continued to grow. The administration of the Home and Community Based Medicaid Waiver program by the U.S. Department of Health and Human Services Centers for Medicare and Medicaid Services (CMS) requires to states to monitor and provide assurances of the quality and performance of their support systems. Federal expectations included in the revised formal Waiver application require that each state agency explain the operational features of its planned HCBS Medicaid Waiver program in detail. States also describe the components of the comprehensive quality management strategy that they intend to employ to monitor and improve the accessibility and quality of services offered to Waiver participants and to demonstrate that required Waiver assurances are being met. Through this approach, CMS has information allowing assessment of a state´s quality management strategy based on the presence or absence of objective, measurable data demonstrating that the state is actively managing the quality of Medicaid Waiver services and supports through a continuous process of discovery, remediation, and improvement. To respond to the emphasis on performance and outcome data, CMS has recommended that states consider the use of standardized tools, such as the National Core Indicators, to gather and analyze information on Waiver participants´ outcomes and satisfaction with the services they receive. The changing expectations on the part of CMS regarding quality monitoring and performance measurement parallel gubernatorial and legislative initiatives in several states aimed at increasing government accountability, improving service outcomes, and strengthening program responsiveness

As PRO-PMs, these measures are designed to assist states in moving beyond compliance to assess service quality, in line with the minimum standards of the federal regulations. The full measure set is designed for states to measure their performance in the HCBS Domains address by the NQF 2016 report and establish goals and benchmarks for performance and leading states to focus on quality improvement, a step beyond compliance.

The 14 measures being submitted collectively addresses 4 of the measurement domains identified by the national, multi-stakeholder Committee on HCBS convened by NQF as areas where “performance measures are needed to drive systems change, tie performance to outcomes, allow consumers to make informed choices, and compare the effectiveness of different models of HCBS and of HCBS versus institutional services.” (NQF, 2016, p. 2).

HCBS Domain: Person-Centered Planning (PCP) and Coordination (Measures #PCP-1 – #PCP-5)

CMS final rules on HCBS Waiver Requirements, published in 2014, promulgated expectations that waiver states establish a Person Centered Planning Process and the Person Centered Plan within a transition period now extended until 2023. The 5 measures in the PCP Domain submitted for NQF endorsement have been developed in response to state DD agencies’ requests for measures which can demonstrate the required compliance with both the person centered planning process, and the HCBS Settings requirements. During annual meetings of the NASDDDS members (State Directors of Developmental Disabilities Services), requests were made to develop and /or adapt existing measures to assist states in determining their performance related to these new rules.

HCBS Domain: Community Inclusion (Measures #CI-1 – #CI-4)

The Americans with Disabilities Act (ADA) of 1990, and subsequent legal actions such as the 1999 Supreme Court decision in the landmark Olmstead case, have established full integration with society as a basic human right for individuals with disabilities. Thus, the adequacy and quality of state DD services depends on the extent to which they support their consumers’ community inclusion. The 4 measures in the Community Inclusion Domain submitted for NQF endorsement measure the success with which state service systems support individuals with IDD in participating in the activities outside their home and integrating with their communities. Given that community inclusion was identified by the NQF Committee on HCBS as one of the key outcomes expected of a service system, these measures are also vital in comparing competing HCBS models.

HCBS Domain: Choice and Control (Measures #CC-1 – #CC-4)

The main purpose of HCBS is to promote and maintain the wellbeing and quality of life of people with functional limitations. Ability to make life choices such as where and with whom to live, services they receive, and how those services are delivered, have been shown to improve the quality of life among people with disabilities (Willis, Grace, & Roy, 2008; Stancliffe et al., 2011). Therefore, measures that assess the level of choice and control that individuals with IDD have over their life and their services are crucial performance measures for state service systems. Monitoring changes in their measures of choice and control will help state service systems achieve continuous quality improvement. The measures are also vital in comparing and improving HCBS systems.

Stancliffe, R.J., Lakin, K.C., Larson, S., Engler, J., Taub, S., & Fortune, J. (2011). Choice of living arrangements. Journal of Intellectual Disability Research, 55(8), 746-762.

Willis, D., Grace, P. J., & Roy, C. (2008). A central unifying focus for the discipline: Facilitating humanization, meaning, choice, quality of life, and healing in living and dying. Advances in Nursing Science, 31(1), 28-40.

HCBS Domain: Human and Legal Rights (Measure #HLR-1)

Within this domain, we are submitting a measure of the respect for one's privacy. This is an important component of basic human rights and dignity that all members of society are entitled to. The Convention on the Rights of Persons with Disabilities promulgates the protection of basic human rights for people with disabilities. Therefore, the extent to which individuals with IDD report that their personal space is respected is an indicator of the ability of an HCBS system to protect its consumers' basic human rights and dignity. It is, therefore, a crucial measure of service quality. Monitoring changes in their scores on this measure will enable state service systems to make the necessary changes to their services to avoid intrusiveness and instead facilitate privacy and dignity among individuals with disabilities.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis*. (****This is required for maintenance of endorsement****. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

Please see tab "2a2.3 and 1b.2" in attached SuppTable_Measures_210420_508.xlsx

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

n/a

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(****This is required for maintenance of endorsement****. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

Please see tab 1b.4 in attached SuppTable_Measures_210420_508.xlsx for disparities data from the measures by population groups.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4

n/a

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, **as specified**, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. Subject/Topic Area *(check all the areas that apply):*

De.6. Non-Condition Specific*(check all the areas that apply):*

De.7. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*

S.1. Measure-specific Web Page *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

Measure-specific web pages are under development.

S.2a. If this is an eMeasure, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is not an eMeasure **Attachment:**

S.2b. Data Dictionary, Code Table, or Value Sets *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*

Attachment **:** SuppTable_Measures_210420_508.xlsx

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Attachment **:** TOOL_2018-19_IPS_final.pdf

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Patient

S.3.1. For maintenance of endorsement: Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

S.3.2. For maintenance of endorsement, please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

IF an OUTCOME MEASURE*, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

The NCI for ID/DD HCBS Measures use values between 0 and 1 as the scores. Typically, the numerator is the number of respondents who selected the most positive response category (e.g. "yes", "always"). The attached file SuppTable_Measures_210420_508.xlsx lists what constituted the most positive response categories for each measure item, as well as other detailed information as relevant for S.2b.

S.5. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

IF an OUTCOME MEASURE, *describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

The attached file SuppTable_Measures_210420_508.xlsx lists detailed information as relevant for S.2b.

Numerators:

- Paid Community Job Goal: The number of respondents who report that community employment is a goal in person´s service plan
- Person-Centered Goals: The number of respondents who report their service plan includes things that are important to them
- ADL Goal: The number of respondents in whose service plan there is a goal to increase independence or improve functional skill performance in activities of daily living (ADLs)
- Lifelong Learning: The number of respondents who report they are supported to learn new things
- Satisfaction with Community Inclusion Scale: This is a multi-item measure, therefore it does not have a simple form for the numerator and denominator
- Social Connectedness: The number of respondents who report that they do not feel lonely often
- Has Friends: The number of respondents who report that they have friends who are not staff or family members
- Transportation Availability Scale: This is a multi-item measure, therefore it does not have a simple form for the numerator and denominator
- Community Inclusion Scale: This is a multi-item measure, therefore it does not have a simple form for the numerator and denominator
- Chose Staff: The number of respondents who report they chose or were aware they could request to change their staff
- Chose Case Manager: The number of respondents who report they could change their case manager/service coordinator
- Can Stay Home When Others Leave: The number of respondents who report they can stay home if they choose when others in their house/home go somewhere
- Life Decisions Scale: This is a multi-item measure, therefore it does not have a simple form for the numerator and denominator
- Respect for Personal Space Scale: This is a multi-item measure, therefore it does not have a simple form for the numerator and denominator

S.6. Denominator Statement *(Brief, narrative description of the target population being measured)*

For each measure, the denominator is the number of respondents (adult recipients of state developmental disabilities services) who provided valid answers to the respective survey question, except those that meet the exclusion criteria (see S.8. below for details).

If the denominator for a state is fewer than 20, the measure score is censored to protect the confidentiality of respondents.

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

***IF an OUTCOME MEASURE**, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

The NCI IPS consists of two main sections, denoted by Roman numerals I and II. Section I of the survey contains questions about personal experiences and therefore may only be answered by the individual receiving developmental disabilities services. Section II of the survey---featuring questions about topics such as community involvement, choices, rights, and access to services—allows for responses from a "proxy," defined as a person who knows the individual well (such as a family member or friend).

Generally speaking, the denominators are the numbers of respondents who are eligible to respond and gave a valid response. Specifically:

#PCP-1: The number of respondents with a valid Section I, who reported that they do not have a job and would like a paid job in the community

#PCP-2: The number of respondents with a valid Section I

#PCP-3: The number of respondents with a valid Section I, who indicated "yes" to the question about desire to increase independence in ADL.

#PCP-4: The number of respondents with a valid Section I

#PCP-5: This is a multi-item measure, therefore it does not have a simple form for the numerator and denominator

#CI-1: Social Connectedness: The number of respondents with a valid Section I

#CI-2: Has Friends: The number of respondents with a valid Section I

#CI-3: This is a multi-item measure, therefore it does not have a simple form for the numerator and denominator

#CI-4: This is a multi-item measure, therefore it does not have a simple form for the numerator and denominator

#CC-1: The number of respondents with a valid Section II

#CC-2: The number of respondents with a valid Section II

#CC-3 The number of respondents with a valid Section I

#CC-4: This is a multi-item measure, therefore it does not have a simple form for the numerator and denominator

#HLR-1: This is a multi-item measure, therefore it does not have a simple form for the numerator and denominator

Exclusion criteria apply. Please see S.8. and S.9. for more details.

S.8. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*

At the end of Section I, the surveyor assesses whether the respondent appears to understand at least one question and answers in a cohesive manner. This assessment is the only subjective process in the exclusion determination process, but it is not done on an arbitrary or state-by-state basis. Rather, it is based on a protocol, included in the survey manual and reviewed during surveyor trainings, that apply uniformly to all surveyors across different participating states. The protocol is straightforward—the section must be marked "valid" if at least one question in the section was answered in a manner that the basic level of comprehension was shown, and a clear response given either verbally (e.g. yes/no) or non-verbally (nodding/shaking head). NCI and participating states routinely conduct surveyor training and surveyor shadowing and reviewing processes that ensure, among other things, that surveyors are applying this assessment (whether or not Section I was valid) strictly based on the protocol. If the surveyor´s assessment is that Section I is not valid, the respondent´s Section I data are flagged for exclusion from the numerators and denominators. However, the individual is not removed from the dataset.

If Section I data are excluded, Section II data are flagged for exclusion from the numerators and denominators -unless- a proxy respondent was used in Section II. If the respondent or proxy did not answer any questions in Section II, the survey is removed from the denominators of Section II items.

Responses are excluded from numerators and denominators for Section I items if:

(a) The surveyor indicated that the respondent did not give consistent and valid responses, or

(b) All questions in Section I were left blank, or marked "not applicable" or "don't know".

Responses are excluded from numerators and denominators for Section II items if:

(a) the individual receiving supports was marked as the sole respondent to all questions in Section II but Section I was deemed invalid, or

(b) All questions in Section II were left blank, or marked "not applicable" or "don't know".

For each measure item, missing responses and responses indicating "not applicable" or "don't know" were excluded from denominators. The distribution of exclusions among states is shown in Testing Attachment 2b2.2. Please see S.9. for more details on denominator exclusions.

S.9. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

In general, missing responses and responses indicating "not applicable" or "don't know" were excluded from denominators. Denominator exclusions for each measure:

- Paid Community Job Goal: Respondents with an invalid Section I (as defined in S.8.), and those who responded, "not applicable" or "don´t know" to the survey question "Would you like to have a job in the community?" are excluded
- Person-Centered Goals: Respondents with an invalid Section I are excluded
- ADL Goal: Respondents with an invalid Section I, and those who did not indicate "yes" to the question about desire to increase independence in ADL are excluded
- Lifelong Learning: Respondents with an invalid Section I are excluded
- Satisfaction with Community Inclusion Scale: Respondents with an invalid Section I are excluded
- Social Connectedness: Respondents with an invalid Section I are excluded
- Has Friends: Respondents with an invalid Section I are excluded
- Transportation Availability Scale: Respondents with an invalid Section I are excluded
- Community Inclusion Scale: Respondents with an invalid Section II are excluded
- Chose Staff: Respondents with an invalid Section II are excluded
- Chose Case Manager: Respondents with an invalid Section II are excluded
- Can Stay Home When Others Leave: Respondents with an invalid Section I are excluded
- Life Decisions Scale: Respondents with an invalid Section II are excluded
- Respect for Personal Space Scale: Respondents with an invalid Section I are excluded

There are no pre-screening procedures prior to the survey. Participation is voluntary, and individual surveys are de-identified. Exclusion of responses occurs at the time of data analysis by HSRI, based on the criteria described above. There is no threshold of number of answers to be met for a "complete" survey.

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate –*

Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)

Risk-adjusted Life Decisions and Community Inclusion Scales, are further stratified by 5 residential setting categories:

category #1 - Intermediate Care Facility for Individuals with Intellectual Disabilities (ICF/IID), nursing facility, or other institutional setting

#2 - Group residential setting (e.g., group home)

#3 - Own home or apartment

#4 - Parents' or relatives' home

#5 - Foster care or host home

There are both conceptual/policy and empirical reasons for this stratification. Conceptually, the need for types and mixes of HCBS supports vary by residential setting, impacting the interpretation and program/policy implications of outcomes. Providing scores for each residential setting separately provides states with meaningful information about the outcomes of these different service/support strategies, offering detailed, actionable recommendations for improvement. Further, risk-adjusted measures significantly vary by residential setting, providing empirical support for the informational value of reporting these measures separately for the 5 settings.

The constructed variable res_type5 was used as the stratification variable. Res_type5 is recoded from background information (administrative records) variable TYPEHOME18, Type of Residence:

The included response TYPEHOME18 categories were:

res_type5 category #1 - ICF/IID, nursing facility or other institutional setting:

1. ICF/IID, 4-6 residents with disabilities
2. ICF/IID, 7-15 residents with disabilities
3. ICF/IID, 16 or more residents with disabilities
4. Nursing facility
5. Other specialized institutional facility

res_type5 category #2 - Group residential setting

6. Group living setting, 2-3 people with disabilities
7. Group living setting, 4-6 people with disabilities
8. Group living setting, 7-15 people with disabilities

res_type5 category #3 - Own home or apartment

9. Lives in own home or apartment; may be owned or rented, or may be sharing with roommate(s) or spouse

res_type5 category #4 - Parent/relative's home

10. Parent/relative's home (may include paid services to family for residential supports)

res_type5 category #5 - Foster or host home

11. Foster care or host home (round-the-clock services provided in a single-family residence where two or more people with a disability live with a person or family who furnishes services)
12. Foster care or host home (round-the-clock services provided in a single-family residence where only one person with a disability lives with a person or family who furnishes services—sometimes called shared living) Other

The TYPEHOME18 categories excluded from res_type5 were:

13. Homeless or crisis bed placement

14. Other (specify):_____

99. Don't know

S.11. Risk Adjustment Type (Select type. Provide specifications for risk stratification in measure testing attachment)

Other

If other: Statistical risk model and stratification

S.12. Type of score:

Rate/proportion

If other:

S.13. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Higher score

S.14. Calculation Algorithm/Measure Logic (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

Please see attached file SuppTable_Measures_210420_508.xlsx for details.

S.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

Each state is instructed to construct a sample frame of adults (18 and over) who are receiving at least one publicly funded ID/DD service in addition to case management. Based on this sample frame and the assumption of a middle response distribution (50%), each state is recommended to have a sample size that will support both (1) a 95% confidence level, and (2) a 5% margin of error. States are allowed to design their own stratifying strategy (e.g. stratifying by Medicaid waiver funding types), as long as they provide the information needed for weighting.

S.16. Survey/Patient-reported data *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

Each state is instructed to construct a sample frame of adults (18 and over) who are receiving at least one publicly funded ID/DD service in addition to case management. Based on this sample frame and the assumption of a middle response distribution (50%), each state is recommended to have a sample size that will support both (1) a 95% confidence level, and (2) a ±5% margin of error.

There are no prescreening procedures. States may approach every sampled individual or stop the recruitment process when their target sample size is reached. Because of these state differences in recruitment processes, no universal response rates are calculated.

Exclusion of responses occurs at the time of data analysis by HSRI, where those without any valid Section I or Section II responses would be considered "incomplete" and excluded from the denominators. The "completeness" of survey is not based on the total number of answers.

S.17. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.18.

Instrument-Based Data

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*

IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

NCI IPS data are collected using the copyrighted survey tools. Up until the 2018-19 survey cycle, the only mode of data collection was face-to-face, in-person survey. Due to the COVID-19 pandemic, remote surveying (via video conferencing) were allowed when following appropriate protocols. NCI IPS is generally administered in English or Spanish.

S.19. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

Available in attached appendix at A.1

S.20. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Population : Regional and State

S.21. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Other

If other: State Home- and Community-Based Services (HCBS) settings

S.22. COMPOSITE Performance Measure - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

2. Validity – See attached Measure Testing Submission Form

NQF_testing_attachment_NCI_210419_508.docx

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

2.2 For maintenance of endorsement

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:**
Measure Title: National Core Indicators for ID/DD Home- and Community-Based Services Measures
Date of Submission: **4/9/2021**

Type of Measure:

Measure	Measure (continued)
☒ Outcome (*including PRO-PM*)	☐ Composite – *STOP – use composite testing form*
☐ Intermediate Clinical Outcome	☐ Cost/resource
☐ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	*

*cell intentionally left blank

NOTE FROM THE MEASURE DEVELOPERS: THE RESPONSES IN BLUE WERE IN THE ORIGINAL TESTING ATTACHMENT SUBMITTED WITH OUR INTENT TO SUBMIT APPLICATION. THE TEXT IN RED WERE SUBSEQUENTLY ADDED IN RESPONSE TO FEEDBACK FROM THE SCIENTIFIC METHODS PANEL.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. ***If there are differences by aspect of testing***,*(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.*

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for* ***all*** *the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.17***)	Measure Tested with Data From:
☐ abstracted from paper record	☐ abstracted from paper record
☐ claims	☐ claims
☐ registry	☐ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: National Core Indicators® In-Person Survey (IPS) Data*	☒ other: National Core Indicators® In-Person Survey (IPS) Data*

Note: Measures described in this document are derived from the National Core Indicators® (NCI®) In-Person Survey (IPS). NCI was developed in 1997 by the National Association of State Directors of Developmental Disabilities Services (NASDDDS), in collaboration with Human Services Research Institute (HSRI). This effort strives to provide states with valid and reliable tools to help improve system performance and better serve people with intellectual and developmental disabilities and their families. Currently, 46 states and the District of Columbia are members of the program, although not every member state conducts the survey every year Reliability and validity testing for NCI survey items has been conducted on an ongoing basis since the late 1990s.

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

Not applicable.

1.3. What are the dates of the data used in testing? July 2018 – June 2019*

*This is the data wave used for the most recent analysis whose results are reported here. However, reliability and validity testing for NCI In-Person Survey items has been conducted on an ongoing basis since the late 1990s. Some of the results cited here come from earlier analyses. The dates of those results are provided wherever they are referenced.

1.4. What levels of analysis were tested? (*testing must be provided for **all** the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: **(*must be consistent with levels entered in item S.20*)**	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☒ other: State	☒ other: Individual service user and state

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

The measured entity for all measures included in this submission is the state. In each member state, the lead agency or accountable entity at the state level that administers services to people with intellectual and developmental disabilities (ID/DD) is responsible for the state's In-Person Survey (IPS) administration, in accordance with NCI's methodological standards. HSRI provides training and technical assistance at all stages of the effort from sampling design through surveyor training and data collection and performs validity checks on the collected data.

IPS data are collected via a direct conversation with a person receiving services from the state's lead agency or accountable entity at the state level that administers services to people with ID/DD. Though eligibility for services varies by state, the population surveyed by the IPS includes individuals with intellectual disabilities and/or developmental disabilities (ID/DD).

The American Association on Intellectual and Developmental Disabilities defines intellectual disability as follows:

> *Intellectual disability is a disability characterized by significant limitations both in intellectual functioning and in adaptive behavior as expressed in conceptual, social, and practical adaptive skills. This disability originates before the age of 18*[6].

The American Association on Intellectual and Developmental Disabilities defines developmental disabilities as follows:

> *Developmental disabilities are severe chronic disabilities that can be cognitive or physical or both. The disabilities appear before the age of 22 and are likely to be lifelong. Some developmental disabilities are largely physical issues, such as cerebral palsy or epilepsy. Some individuals may have a condition*

---

[6] American Association on Intellectual and Developmental Disabilities (2018) *Frequently Asked Questions on Intellectual Disability and the AAIDD Definition.* Retrieved from **https://tinyurl.com/y4rca4pz**

that includes a physical and intellectual disability, for example Down syndrome or fetal alcohol syndrome[7].

The IPS was designed to collect information about service satisfaction, quality of life and important life outcomes directly from individuals receiving services. As such, the IPS offers valid, reliable, person-centered measures that states use to demonstrate how publicly funded supports are impacting people's lives and to determine where state systems can improve the quality of those supports. However, gathering subjective information and data from people with ID/DD poses unique challenges due to potential intellectual and developmental limitations experienced by the population. The processes for developing NCI IPS administration methods, survey methodology and measure design and revision includes literature review, testing and consultation with members of the target population and their advocates.

The IPS consists of two main sections, denoted by Roman numerals I and II. **Section I** of the survey contains questions which pertain to personal experiences and require subjective responses; this section may only be answered by the individual receiving services. **Section II** of the survey—which consists of objective questions on the individual's involvement in the community, their choices, rights, and their access to services—allows for responses from a "proxy," that is, a person who knows the individual well (such as a family member or friend). Surveyor training ensures that surveyors are able to identify acquiescence (e.g. all yes responses), and indicators of inability to understand and respond to questions. At the end of Section I, the surveyor indicates whether the respondent appeared to understand the questions and answered them in a consistent manner. If the surveyor's response to this question is in the negative, Section I data are excluded from analysis. If Section I data are excluded from analysis based on the surveyor's assessment of inconsistent responses and potential lack of understanding, Section II data are also excluded for this case unless a proxy respondent was used. A third part of IPS data, known as "Background information" or BI, comes from administrative records and is used to characterize the demographics of respondents. In some cases, BI data is used to determine whether a question is relevant for the respondent to answer.

The data analyzed for the most recent testing came from 37 states, listed in Exhibit 1 below. The 2a2.3 tab included in the supplemental tables accompanying this attachment specify the number of states whose data were used for the most recent testing results. Each state is instructed to construct a sample frame of adults (18 and over) who are receiving at least one publicly funded ID/DD service in addition to case management. Based on this sample frame and the assumption of a middle response distribution (50%), each state is recommended to have a sample size that will support both (1) a 95% confidence level, and (2) a ±5% margin of error.[8] Most states sample more than this minimum recommended size to account for refusals and surveys that may be deemed invalid. Some states stratify their samples by factors such as region, program, or funding source. State-specific weights are applied to correct for under- or over-sampling within strata. A standard data validation and cleaning procedure is applied to returned surveys to identify inconsistent responses and responses noted by the surveyor to be inconsistent or the result of the respondent's lack of understanding (see above paragraph for details). It should be noted that not all member states collect all of the data elements required to construct all of the proposed measures, although most states include most of the items in their surveys.

Exhibit 1. Number and Percentage of Valid Surveys by State

---

[7] American Association on Intellectual and Developmental Disabilities (ND) *Frequently Asked Questions on Intellectual Disability.* Retrieved from https://www.aaidd.org/intellectual-disability/definition/faqs-on-intellectual-disability

[8] The calculation of minimum sample size required to satisfy these two criteria is not based on a specific measure. Rather, it is calculated for a hypothetical dichotomous variable with the maximum possible variance (i.e. a mean of 0.5). All of the 14 measures being submitted are based on dichotomous variables and will therefore have a variance less than or equal to the one assumed for this calculation.

State	% Valid Answers to Section I	% Valid Answers to Section II	Number of Valid Surveys (Section I and II) Completed
AL	91.0%	99.7%	378
AR	78.4%	99.4%	510
AZ	71.2%	100.0%	413
CO	95.3%	97.9%	712
CT	71.4%	96.6%	391
DE	71.1%	95.6%	342
FL	70.9%	98.5%	867
GA	68.9%	99.6%	479
HI	45.2%	99.4%	363
IN	71.2%	99.1%	742
KS	68.5%	98.8%	408
KY	62.7%	100.0%	429
ME	83.8%	95.0%	400
MI	63.1%	99.7%	658
MN	73.9%	94.8%	2140
MO	72.2%	100.0%	403
NC	57.1%	99.8%	651
NE	74.4%	99.5%	418
NH	77.0%	98.0%	400
NJ	65.6%	99.6%	514
NV	84.8%	98.3%	408
NY	54.2%	94.4%	448
OH	74.2%	99.0%	732
OK	69.0%	99.5%	400
OR	78.9%	99.8%	421
PA	73.8%	98.9%	710
RI	80.7%	98.9%	358
SC	96.2%	100.0%	501
SD	88.0%	93.4%	351
TN	81.8%	100.0%	406
TX	59.8%	100.0%	2388
UT	63.4%	98.7%	372

State	% Valid Answers to Section I	% Valid Answers to Section II	Number of Valid Surveys (Section I and II) Completed
VA	73.9%	99.8%	807
VT	74.3%	98.6%	346
WA	72.1%	99.3%	441
WI	67.2%	94.4%	961
WY	80.4%	92.4%	341
Total	**71.6%**	**98.2%**	**22,009**

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

The characteristics or the 22,009 valid cases are as follows (numbers based on weighted cases, information obtained from participant records, unless otherwise noted).

- **Proxy Used in Any *Section II* Items?** (source: surveyor report): 24.7% yes, 75.3% no;
- **Age:** 6.0% 18 – 22, 32.2% 23 – 34, 36.6% 35 – 54, 23.1% 55 – 74, 2.1% 75 or older;
- **Gender:** 60.5% male, 39.4% female, 0.1% other;
- **Race:** 72.8% Non-Hispanic White; 15.1% Non-Hispanic Black; 6.7% Non-Hispanic all other races, 5.4% Hispanic/Latinx;
- **Residential Setting**: 18.7% own home, 36.0% parents'/relatives' home; 7.9% foster/host home, 31.2% group residential setting, 6.2% institutional setting;
- **Type of Community of Zip Code Area** (source: Department of Agriculture's Rural-Urban Commuting Area codes)**:** 70.7% metropolitan, 16.5% micropolitan, 8.7% small town, 4.1% rural;
- **Median Income of Zip Code Area** (in $1,000; source: American Community Survey 5-year estimates)**:** 4.9% 0 – 29, 19.2% 30 – 39, 48.2% 40 – 59, 25.6% 60 – 99, 2.0% 100 or higher;
- **Funding Source(s) for Supports:** 6.1% ICF/IID, 83.3% Medicaid HCBS Waiver, 20.5% Medicaid State Plan, 1.0% Exclusively supported by state funds (no Medicaid services beyond health care);
- **Medicare Recipient?** 57.2% yes, 42.8% no;
- **Legal Status:** 45.9% not in guardianship relationship, 54.1% under full or limited guardianship;
- **Diagnosed with Intellectual Disability (ID)?** 91.6% yes, 8.4% no;
- **Level of ID Among Those with ID Diagnosis:** 41.5% mild, 30.7% moderate, 18.6% severe or profound; 9.2% level unspecified;
- **Primary Language:** 97.5% English, 2.5% other;*
- **Preferred Means of Communication:** 82.6% spoken; 17.4% non-spoken;
- **Mobility Status:** 79.9% ambulatory without wheelchair or aid, 13.4% ambulatory with wheelchair or aid, 6.8% non-ambulatory;
- **Employment:** 27.6% with paid job in the community; 72.4% no paid job in the community.
- **Overall Health Status** (based on self or proxy responses): 18.8% excellent, 78.0% very or fairly good, 3.2% poor.

---

*There is a standard Spanish version of the instrument.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

The data described above was used for all testing.

1.8 What were the social risk factors that were available and analyzed? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

We include, in this section, all factors considered for risk adjustment, without distinguishing between clinical and social risk factors, since in the context of HCBS outcomes, the distinction is not as clear-cut as it would be for clinical measures.

The analysis conducted to select risk adjustment factors considered the following variables:

From participant records: Age, gender, race, marital status, type of residence, length of stay at current residence, level of intellectual disability, presence of other disabilities (hearing or vision loss), physical and mental health diagnoses, presence of disruptive and destructive behaviors, ambulatory ability, preferred means of expression, primary language.

Surveyor-reported: Whether or not a proxy was used for at least one survey item.

Self- or proxy-reported: Overall health status.

Through linkage to external data (see 1.6 for data sources)**:** Median income and urbanicity/rurality of the zip code area.

2a2. RELIABILITY TESTING

Note*: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)
2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Reliability Tests on Critical Data Elements

Measure names, numbers, and critical elements used in calculating the 14 measures described in this submission are described in detail in the **"Measure Descriptions" tab of the Supplemental Tables** attachment. The critical elements for all but two of the 14 measures require responses based on an interview. The two exceptions are whether or not the service plan includes an employment goal (used to construct Community Job Goal, #PCP-1) and whether or not the plan includes a goal to improve functional skill performance in ADLs (used to construct ADL Goal, #PCP-3); the data for these two items come from administrative records. Seven of the 14 measures use responses from a single survey item and the remaining 7 require multiple items for their calculation; 5 of those 7 are multi-item scales that combine responses to multiple items in their calculation (PCP-5, CI-3, CI-4, CC-4, and HLR-1) and the remaining two (PCP-1 and PCP-3) rely on two items each for inclusion in the numerator and denominator.

Since its inception in the late 1990s, the NCI In-Person Survey has undergone multiple **inter-rater reliability (IRR) studies**. IRR analysis consists of techniques to assess the level of agreement among multiple observers of the same event. In this case, multiple individuals recorded the participants' responses to the same questions through in-person or video "shadow interviews." The results of the IRR studies were summarized by calculating **percent agreement** among raters and **Cronbach's Kappa**. Kappa ranges between 0 (no agreement between raters beyond that which would be expected to occur purely by chance) and 1 (perfect agreement between raters).

Reliability and validity of the data that come from administrative records in response to the Background Information section of the instrument were tested by having multiple individuals abstract the information using the same records.

To ensure that survey questions are understood, a **cognitive test** (cog test for short) was conducted specifically for the critical elements of the measures described in this document. Ten IDD service users from member states participated in the study. The test instrument consisted of 35 survey items. Cognitive tests are surveys conducted with a group of individuals from the target population that include the survey items being tested *plus* follow-up questions to the respondent and the surveyor about how each question was understood and interpreted by the respondent. The results of the pretest provide information about how well a typical respondent understands the questions, whether or not the intent of each question is interpreted similarly by different respondents, and whether or not a typical respondent is able to provide consistent and relevant responses to the set of questions being tested.

The multi-item scales were constructed based on the results of **principal components analysis (PCA).** This is an exploratory factor analysis conducted on multiple candidate items for a scale, to select the group of items that best represent a single cohesive factor, by assessing the extent to which each candidate item relates to the overall factor.

Once the group of items that best represent the five multi-item scales (CC-4, HLR-1, CI-3, CI-4, and PCP-5) were identified through exploratory factor analysis, **confirmatory factor analysis** (CFA) was conducted to test the overall factor structure of the five multi-item scales. This is a multivariate statistical procedure that is used to simultaneously test the appropriateness of all five scales in representing the observed data.

We calculated **Cronbach's Alpha** and Spearman-Brown Coefficient to measure the internal consistency of the multi-item scales. Both statistics are measures of the extent to which the constituent items of a scale are related to each other. The higher the Alpha or the Spearman-Brown coefficient, the more internally consistent is the scale. In addition, we examined the **Corrected Item-Total Correlation Coefficients**. These coefficients measure the correlation between each item in the scale and the calculated scale. The correction involves removing the item from the scale to avoid artificially inflating the correlation. High and positive values of these corrected correlation coefficients indicate better internal consistency.

A table identifying the type of reliability testing reported for each of the 14 measures is provided in the "Reliability Test by Measure" tab of the Supplemental Tables attachment.

Reliability Tests on Performance Measure Scores

We conducted **analyses of variance** (ANOVA) on all 14 measures being submitted, to test whether the between-state variance is significantly larger than the within-state variance. The results of these analyses assess the success with which the measures distinguish among the states (measured entities). In the context of measuring the performance of a state's service system, variation among individual service users within a state would be considered "noise" while variation among states' scores would be considered the "signal" (i.e., usable information). In addition to the ANOVA F-statistic that tests the within- and between-state variation difference, we calculated **inter-unit reliability** statistics (IUR) for each measure. Although similar to the F-statistic in purpose and function — it is a transformation of F calculated as (F-1)/F — IUR has a more intuitive interpretation. It represents the proportion of between-state variance (i.e. the signal) left over after removing the within-state variance (i.e., the "noise"). In that sense, it is a measure of the measure's signal-to-noise ratio.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Critical Data Element Test Results

The NCI In-Person Survey (IPS, previously known as the Adult Consumer Survey) was designed with input from a Project Advisory Committee with extensive experience in instrument development, and knowledge of methods used to measure service system outcomes. A comprehensive literature review of outcome-based research and evaluation also informed the process.[9] The tool has undergone rigorous testing, both during the initial piloting and after significant revisions were made. Selected results from **IRR studies** are as follows:

- In 1997, a pilot test of the NCI tool was conducted with 30 individuals in a member state. Inter-rater reliability testing resulted in 93% average agreement between the raters.
- A 1998, inter-rater reliability study (N=25) resulted in 93% agreement between the raters, and an average Kappa score of 0.794.
- A 1999 reliability test (N=27) found 92% agreement between raters.
- In 2008, the survey underwent some revisions, and a pilot test was conducted with 16 individuals in a member state. Inter-rater reliability tests of this sample resulted in an average Kappa score of 0.90 across pairs of raters.

In 2010, the NCI Team conducted a study where 20 interviews were conducted by six trained surveyors from a member state and “shadowed” by NCI team members who independently recorded the responses. The Kappa scores for the six interviewers ranged from 0.82 to 0.95, with an average of 0.89. Average percent agreement for the overall survey was 80%. Average percent agreement for survey sub-sections were as follows:

Survey Section	Avg. Pct. Agreement	Relevant Measures
Work/Daytime Activities	75%	PCP-1, PCP-3
Home	68%	CC-3
Friends and Family	71%	CI-1, CI-2
Satisfaction with Services	71%	PCP-2
Community Inclusion	95%	PCP-5, CI-4
Choices	90%	CC-1, CC-2, CC-4
Rights	91%	HLR-1
Access to Needed Services	90%	PCP-4, CI-3

It is important to note that most of the data elements for the 14 measures being submitted come from the Community Inclusion, Choices, and Rights sections of the instrument with 90% agreement on average.

The IRR study focusing on the reliability of background information from state records yielded 90% or higher agreement among different coders using the same data sources.

Results of the **principle components analysis** were as follows:

- The 5 items comprising the *Life Decisions Scale* constituted a single factor that explained 45% of the variance with component loadings ranging from 0.521 to 0.759.
- The 3 items comprising the *Respect for Personal Space Scale* constituted a single factor that explained 44% of the variance with component loadings ranging from 0.490 to 0.746.
- The 2 items comprising the *Transportation Availability Scale* constituted a single factor that explained 71% of the variance with both components loading at 0.842.
- The 4 items comprising the *Community Inclusion Scale* constituted a single factor that explained 44% of the variance with component loadings ranging from 0.583 to 0.720.

---

[9] Smith, G. & Ashbaugh, J. (2001). National Core Indicators Project: Phase II Consumer Survey Technical Report. Retrievable from: http://www.nationalcoreindicators.org

- The 5 items comprising the *Satisfaction with Community Inclusion Scale* constituted a single factor that explained 46% of the variance with component loadings ranging from 0.583 to 0.763.

Cronbach's Alpha and **Spearman-Brown Coefficient** (SBC) for the multi-item scales, based on the 2018-2019 data are as follows:

- CC-4. Life Decisions Scale: Alpha=0.673, SBC=0.666
- HLR-1. Respect for Personal Space Scale: Alpha=0.349, SBC=0.234
- CI-3. Transportation Availability Scale: Alpha=0.559, SBC=0.593
- CI-4. Community Inclusion Scale: Alpha=0.489, SBC=0.460
- PCP-5. Satisfaction with Community Inclusion Scale: Alpha=0.694, SBC=0.727

Please note that the Transportation Availability Scale (CI-3) is composed of two data elements while the rest of the scales have at least three elements. For two-item scales, the Spearman-Brown coefficient is a more appropriate reliability measure than Cronbach's Alpha,[10] and was used to assess reliability for CI-3.

The corrected item-total correlation coefficients for the scales are displayed in Exhibit 2 below. These coefficients represent the correlation between each item within a scale and the scale calculated without that item.

Exhibit 2. Corrected item-total correlations (CITC)

Scale	Item	CITC Coefficient
CC-4. Life Decisions Scale	Did you choose the place where you live?	0.508
CC-4. Life Decisions Scale	Did you choose the people you live with?	0.403
CC-4. Life Decisions Scale	Did you choose the place where you work?	0.431
CC-4. Life Decisions Scale	Did you choose your day program or workshop?	0.510
CC-4. Life Decisions Scale	Did you choose your staff	0.310
HLR-1. Respect for Personal Space Scale	Do people let you know before they come into your home?	0.243
HLR-1. Respect for Personal Space Scale	Do people let you know before they come into your bedroom?	0.257
HLR-1. Respect for Personal Space Scale	Do you have a place to be alone in your home?	0.134

---

[10] Eisinga, R., Grotenhuis, M.t., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach or Spearman-Brown? *International Journal of Public Health, 58*, 637–642. doi: 10.1007/s00038-012-0416-3

Scale	Item	CITC Coefficient
CI-3. Transportation Availability Scale	Do you have a way to get to places you need to go?	0.422
CI-3. Transportation Availability Scale	Are you able to get to places when you want to do something fun?	0.422
CI-4. Community Inclusion Scale	How many times did you go shopping in the past month?	0.387
CI-4. Community Inclusion Scale	How many times did you go out on errands or appointments in the past month?	0.292
CI-4. Community Inclusion Scale	How many times did you go out for entertainment in the past month?	0.337
CI-4. Community Inclusion Scale	How many times did you go to a restaurant or coffee shop in the past month?	0.403
PCP-5. Satisfaction with Community Inclusion Scale	Would you like to go shopping more, less, or about the same amount as now?	0.472
PCP-5. Satisfaction with Community Inclusion Scale	Would you like to go out for entertainment more, less, or the same amount as now?	0.551
PCP-5. Satisfaction with Community Inclusion Scale	Would you like to go out to a restaurant or coffee shop more, less, or the same amount as now?	0.512
PCP-5. Satisfaction with Community Inclusion Scale	Would you like to go to religious services or spiritual practices more, less, or the same amount as now?	0.391
PCP-5. Satisfaction with Community Inclusion Scale	Do you want to be a part of more groups in your community?	0.376

We conducted **confirmatory factor analysis** to test the overall factor structure of the five multi-item scales (CC-4, HLR-1, CI-3, CI-4, and PCP-5). The model fit statistics from the estimated structural equations model (SEM) are as follows:

- Tucker Lewis index (TLI) = 0.924
- Comparative Fit Index (CFI) = 0.942
- Root Mean Square Error of Approximation (RMSEA) = 0.026

The results of the **cognitive pretest** indicated that overall, the questions did not pose undue cognitive burden for respondents. The following are summary quantitative results from the 35 test questions administered to 10 respondents:

- The number of valid responses to the items ranged between 8 and 10 with an overall average of 9 over all 35 items.
- After most questions, respondents were asked whether they found the question easy or hard to answer, on a 4-point Likert scale ranging from 0 (very easy) to 3 (very hard). Item averages ranged

between 0.5 and 1.5, with an overall average of 1.1 over all questions for which the difficulty assessment was solicited.

- After each question, surveyors indicated whether they had to repeat or clarify the question. Number of repeats per question ranged from 0 to 2, with an overall average of 0.30 repeats over all items. Number of clarifications per question ranged from 0 to 4 with an overall average of 0.9 clarifications over all items.
- 47% of the questions had no repetition or clarification request from any of the respondents.

Qualitative analysis of responses to open-ended items (e.g. "What came to your mind when you first heard the word….?") also indicated that the items were understood well and consistently by the respondents. One insight from this analysis indicated that the easy/hard assessment in the quantitative results overestimates cognitive burden. Multiple responses to the open-ended question, "Why did you find the question hard to answer" pointed to emotional/sentimental difficulties, not cognitive ones. For example, the question on ability to change the case manager was reported as being hard to answer because the participant liked and would never want to change the case manager. Likewise, the loneliness item was reported as being hard to answer because it brought up a topic that was painful.

Performance Measure Score Test Results

ANOVA analysis found that between-state variation is significantly larger than within-state variation for each of the 14 measures (p<0.001). The IUR ranged between 0.753 and 0.984. Detailed distributional properties, sample sizes, and IUR for each measure is provided in the Supplemental Tables workbook, tab "2a2.3 and 1b.2".

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The multiple **inter-rater reliability** (IRR) tests conducted on NCI In-Person Survey items show high levels of agreement among raters. Overall agreement that these studies yielded ranged between 80% and 93%. The sections of the instrument that contain the critical elements of the 14 measures described here (Community Inclusion, Choices, Rights, and Access to Needed Services) all have average percent agreement scores above 90%. These results indicate that the critical elements are coded in a dependable and consistent manner by the surveyors, thus providing evidence for the reliability of the data. These positive results may be attributed, at least in part, to the detailed manuals and surveyor training curricula developed for NCI surveys and frequent "shadow interviews" and troubleshooting sessions provided by NCI team members.

The results of the **principal components analysis** suggest an acceptable factor structure and internal consistency for the multi-item scales. For example, a minimum factor loading of 0.4 is recommended by multiple scholars to conclude that a specific component "fits" into a scale.[11] None of the component loadings of the five scales was below that threshold and most were in the 0.6 – 0.7 range, suggesting that the component items fit the overall scale well. The components explained approximately 40% of the shared variance in all five scales. Percent explained variance may be interpreted as the proportion of "information" contained by the different components that is retained by combining the components into a scale. These are acceptable preliminary results. The scales will be subjected to confirmatory factor analysis to further investigate the factor structure of the scales.

We conducted **confirmatory factor analysis** to test the overall factor structure of the five multi-item scales (CC-4, HLR-1, CI-3, CI-4, and PCP-5). The results of the **confirmatory factor analysis** indicate that the measurement model represented by the five multi-item scales fits the data reasonably well:

- Tucker Lewis index (TLI) = 0.924 (≥ 0.90 indicates acceptable fit to the data)

---

[11] Raubenheimer, J. (2004). An item selection procedure to maximize scale reliability and validity. *SA Journal of Industrial Psychology, 30*(4), 59-64.

- Comparative Fit Index (CFI) = 0.942 (≥ 0.90 indicates acceptable fit)
- Root Mean Square Error of Approximation (RMSEA) = 0.026 (< 0.05 indicates good fit)

For a scale to be meaningful, the items composing it have to have some level of association with each other and with the calculated scale. **Cronbach's alpha** ranges between 0 and 1 and can be interpreted as the level of association among the constituent items in a scale. An alpha of means that the items are completely independent of each other; that is, they do not share any information. An alpha of 1 means that the items perfectly correlate with one another. Although the dividing line between "good" and "bad" alpha scores is somewhat arbitrary, there is some consensus that a score of 0.7 or above indicates high reliability and a score below 0.5 does not qualify as reliable.[12] By these standards, four of the five NCI scales reach or approach good reliability. The sole exception is the *Respect for Personal Space Scale* with an alpha of 0.349. The constituent items of this scale have limited variability, at least partially explaining the relatively low alpha.

Another measure of the reliability of a scale is the association between the calculated scale and each of its items. The corrected item-total correlations CITC provided in Exhibit 2 above provide measures of these associations, "correcting" the scale measure by removing the constituent item to avoid artificially inflating the correlation. A minimum CITC coefficient of 0.4 is considered by many scholars to indicate reliability.[13] Thirteen of the 18 CITC coefficients calculated for the five scales approach (i.e. round up to) or exceed this threshold and 4 scales have at least one element approaching the threshold. Predictably, *Respect for Personal Space Scale* is the sole exception, possibly explained by the low variability of the items.

There are no clearly articulated measures and norms for assessing cognitive pretesting results. The test conducted by the NCI team used several quantitative measures that measure consistent comprehension of the survey items, such as participant-reported level of difficulty and surveyor-reported item repetition and clarification. All of these indicated high overall comprehension. The qualitative analysis results supported these findings and offered further evidence that the ways in which the questions were interpreted by different respondents did not vary substantially.

Interpretation of Performance Measure Score Test Results

The Supplemental Table tab "2a2.3 and 1b.2" shows the distributional properties and inter-unit reliability scores of the measures aggregated to the state level. A minimum IUR of 0.7 is recommended as an acceptable measure of signal-to-noise ratio for a clinical performance measure.[14] All of the measures have an IUR well above that threshold and for some measures (e.g. *CC-1:Chose Staff, CC-4: Life Decisions Scale, PCP-5: Satisfaction with Community Inclusion Scale*), practically the entire variation is between states. In interpreting these numbers, it should be kept in mind that the IUR was developed to measure the clinical performance of physicians and medical facilities. In such cases, the scores are less subject to variability in social determinants of health across units, compared to HCBS measures with states as units. There are also legal and policy differences among states that add to within-state homogeneity, increasing the IUR. These considerations point to the policy-relevance and actionability of these measures. For example, the within-state homogeneity of measures such as the *Life Decisions* or *Satisfaction with Community Inclusion* scales, opens up areas of inquiry into state policies and practices that facilitate or impede personal choice, thus providing important policy guidance.

---

[12] Ware, J.E., Jr, Gandek, B. (1998). Methods for testing data quality, scaling assumptions, and reliability: the IQOLA project approach. *Journal of Clinical Epidemiology, 51*(11):945–952.

[13] Ware, Jr. & Gandek (1998).

[14] Scholle SH, Roski J, Adams JL, Dunn DL, Kerr EA, Dugan DP, Jensen RE. (2008). Benchmarking physician performance: reliability of individual and composite measures. *American Journal of Managed Care, 14*(12):833-8. PMID: 19067500; PMCID: PMC2667340.

It is also important to note that the distributional characteristics and analysis of variance results reported here come from non-adjusted versions of the measures. As discussed below in Section 2b3, risk adjustment and stratification will, to some extent, reduce the between-state variation observed in non-adjusted measures.[15]

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

☒ **Empirical validity testing**

☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Critical Element Level

The NCI In-Person Survey was developed with extensive involvement of the Project Advisory Committee of state officials and other advisors to assure the survey validly represented the established national goals of services for persons with ID/DD. Additionally, the draft survey was reviewed by a focus group of individuals with intellectual disabilities to pre-test the face validity of the questions. Focus group participants highlighted problematic questions, identified words that needed further definition, and suggested alternative ways of phrasing questions. These modifications were incorporated into the final survey.

To assure individual interview validity, interviewers are asked to give formal feedback on every interview conducted. At the end of Section I, there are two questions that ask the interviewer to make a judgment about the individual's comprehension of the questions and consistency of responses. This provides project staff with information on survey questions that were confusing to an interviewee and help to further clarify questions and interviewer instructions each year. These items are also used during analyses to identify respondents the interviewer doubted had understood questions adequately to have responded reliably and validly.

Over the years, multiple published studies investigated the relationships among NCI data elements tested hypotheses about expected associations. The following is selected list of published studies:

Houseworth, J., Stancliffe, R.J., & Tichá, R. (2018). Association of state-level and individual-level factors with choice making of individuals with intellectual and developmental disabilities. *Research in Developmental Disabilities, 83*, 77-90.

> Used multilevel modeling on NCI data to explore the ways in which the NCI choice scales correlated with several conceptually related variables. The results indicate that as hypothesized, At the individual level, milder ID, greater mobility, fewer problem behaviors, answering questions independently, communicating verbally, and living in a non-agency setting, particularly independent settings, were associated with more choice.

Houseworth, J., Stancliffe, R., & Tichá, R. (2019). Examining the National Core Indicators' Potential to Monitor Rights of People with Intellectual and Developmental Disabilities According to the CRPD. *Journal of Policy and Practice in Intellectual Disabilities, 16*(4), 342-351.

> This study was designed to assess empirically the extent to which people with intellectual and developmental disabilities (IDD) exercise certain rights in the United States using the National Core Indicators Adult Consumer Survey (NCI). An additional aim was to assess the impact of guardianship on

---

[15] He, K., Dahlerus, C., Xia, L., Li, Y., & Kalbfleisch, J.D. (2020). The profile inter-unit reliability. *Biometrics, 76*(2), 654-663.

the rights of individuals with IDD. NCI data were analyzed employing factor analysis, multiple analysis of variance, and regression modeling. These approaches allowed us to assess the relationship between guardianship and rights controlling for known covariates (such as level of ID) on outcomes. Results indicate that the NCI contains several items with sound psychometric properties that can assist in measuring certain rights of people with disabilities according to CRPD. Specifically, employment and budgetary agency appear to be areas of rights outlined by the CRPD that the NCI can help measure. Finally, the results indicated that people who have an appointed legal guardian are less likely to be employed and to have less social privacy. This study indicates the NCI has the potential to measure access to CRPD rights by people with IDD.

Lakin, K. C., Doljanac, R., Byun, S., Stancliffe, R. J., Taub, S., & Chiri, G. (2008). Choice making among Medicaid Home and Community-Based Services (HCBS) and ICF/MR recipients in six states. *American Journal on Mental Retardation, 113* (5), 325-342.

The authors used multivariate statistical techniques on NCI data to compare the level of personal choice in life decisions and support services between adults with IDD receiving Medicaid Home and Community Based Services (HCBS) and those receiving Intermediate Care Facility (ICF/MR) services and living in non-family settings. After controlling for level of intellectual disability, medical care needs, mobility, behavioral and psychiatric conditions, and self vs. proxy reporting, they found that choice was more strongly associated with residential setting than whether that setting was HCBS- or ICF/MR-financed.

Mehling, M.H., & Tasse, M.J. (2014). Empirically derived model of social outcomes and predictors for adults with ASD. *Intellectual and Developmental Disabilities, 52*(4), 282-295.

This study used NCI data to derive an empirically validated measurement model for social outcomes and associated constructs for both individuals with Autism Spectrum Disorder (ASD) and individuals with other disabilities. Items consistent with the survey structure of the NCI were selected as initial indicators of the latent constructs Social Relationships, Community Inclusion, and Opportunity for Choice in factor analyses. Results yielded a novel factor structure that is different from the original NCI survey structure. Three factors emerged as a result of these analyses: Personal Control, Social Determination, and Social Participation and Relationships. The factor structure of each of these constructs was consistent although not identical across individuals with ASD and individuals with developmental disabilities other than ASD.

Stancliffe, R. J., Lakin, K. C., Taub, S., Chiri, G., & Byun, S. (2009). Satisfaction and sense of well-being among Medicaid ICF/MR and HCBS recipients in six states. *Intellectual and Developmental Disabilities (formerly Mental Retardation), 47*(2), 63-83.

Self-reported satisfaction and sense of well-being were assessed in a sample of 1,885 adults with intellectual and developmental disabilities receiving Medicaid Home and Community Based Services (HCBS) and Intermediate Care Facility (ICF/MR) services in 6 states. Questions dealt with such topics as loneliness, feeling afraid at home and in one's neighborhood, feeling happy, feeling that staff are nice and polite, and liking one's home and work/day program. Loneliness was the most widespread problem, and there were also small percentages of people who reported negative views in other areas. Few differences were evident by HCBS and ICF/MR status. The findings document consistent benefits of residential support provided in very small settings-with choices of where and with whom to live-and to individuals living with family.

Tichá, R., Lakin, K.C., Larson, S., & Stancliffe, R., Taub, S., Engler, J., Bershadsky, J., & Moseley, C. (2012). Correlates of everyday choice and support-related choice for 8,892 randomly sampled adults with intellectual and developmental disabilities in 19 states. *Intellectual and Developmental Disabilities, 50*(6), 486-504.

This study used NCI data to examine the correlates of choice among adults with IDD. Results indicate that controlling for physical and sensory impairment, age, behavioral support, communication, and state, people in residential settings with 16 or more people had less everyday choice than those in other living arrangements. People with mild and moderate IDD had more control over everyday choices when living in their own homes, whereas people with severe and profound IDD had more control when living in agency homes of 3 or fewer residents. For people of all levels of IDD, institutional settings of 16 or more residents offered the lowest levels of everyday choice. Controlling for the same covariates, individuals with all levels of IDD living in their own homes had significantly more support-related choices than those in any other residential arrangement. Controlling for individual and residential setting characteristics, the state in which sample members lived was notably predictive of support-related choice. Overall, the tested variables accounted for 44% of the variability in everyday choice and 31% in support-related choice.

Tichá, R., Qian, X., Stancliffe, R.J., Larson, S., & Bonardi, A. (2018). Alignment between the Convention on the Rights of Persons with Disabilities and the National Core Indicators. *Journal of Policy and Practice in Intellectual Disabilities, 15*(3), 247-255.

On December 13, 2006, the United Nations adopted the Convention on Rights on Persons with Disabilities (CRPD). An important component of CRPD implementation is a systematic effort to evaluate whether and what extent the rights as stated in the CRPD are being upheld by individual countries. The purpose of this paper was to contribute to a U.S. and international discussion about monitoring the rights of persons with disabilities as codified by the CRPD by examining the potential use of the National Core Indicators Adult Consumer Survey (NCI‐ACS) for this purpose. The NCI‐ACS collects annual quality outcome data from over 13 000 people with intellectual and developmental disabilities in 46 U.S. states. A crosswalk between the CRPD Articles and the 2013–14 NCI‐ACS was conducted to map the survey questions onto the CRPD. Forty‐seven variables from the NCI‐ACS focused on different aspects of disability rights were identified as related to 12 CRPD Articles, suggesting a moderate overlap between the NCI‐ACS and the CRPD.

Performance Measure Score Level

At the state-level, we conducted analyses to test the **criterion validity** of the measures. This analysis assessed the extent to which a state's score on a measure correlates with another factor (generally referred to as the criterion) with which it would be theoretically expected to correlate. A criterion can be any variable for which there is a theoretical or logical reason to expect an association with the measure being tested. If a correlation is found between the measure and the criterion, the result provides evidence that the measure is a valid representation of the concept it is intended to measure.

The criteria we selected for this analysis included factors external to the state's HCBS system (such as socioeconomic factors) as well as measures among the 14-measure set being submitted for endorsement. Our use of criteria from within the measure set are justified by the fact that 14 measures assess different aspects of the state's service system; therefore, correlations among them in the logically expected direction can be considered evidence of criterion validity. In other words, intercorrelations among the 14 measures in the logically expected direction also provide evidence of criterion validity.

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Exhibit 3 below displays the criteria, the logical/theoretical reason for expecting an association, and the results of the statistical test measuring the association. Depending on the distributional properties of the measure, either Pearson's Correlation Coefficients (r) or Spearman's Rank Correlation Coefficient (ρ) were used as the measure of association between two measures. T-test of the difference between means was used to test whether the mean score among states with a given characteristic significantly differed from the mean score of states without that characteristic. Two-tailed tests were used to assess statistical significance. P-values less than 0.05 indicate statistical significance.

Exhibit 3. State-Level Criterion Validity Test Results

Measure Being Tested	Criterion	Relational Hypothesis	Test Results
Community Job Goal (PCP-1)	Proportion of the state's population living in urban areas. [Source: U.S. Census Bureau]	Urban settings provide a broader range of employment opportunities and hence, a larger choice of the types of jobs that are available for people with IDD. Urbanized states would be expected to find it easier to meet individuals' need for employment by including their wish for employment in their service plans.	Percentage of a state's population living in urban areas is positively and significantly correlated with PCP-1. $r = 0.395$ $p = 0.011$
Community Job Goal (PCP-1)	Membership in the State Employment Leadership Network (SELN)	(SELN) was established in 2006 to support state public managers to offer expanded community-based employment options for people with IDD. We would expect SELN member states to have greater capacity and incentives to support the employment needs of their service participants and hence, to score higher on this measure.	Mean PCP-1 score for SELN member states is 0.4146 compared to a mean score of 0.3347 for non-member states. A one-tailed t-test of the difference between means yields a p-value of 0.055. Keeping in mind the low sample size (37 states), this result provides some support for the hypothesis.
Person-Centered Goals (PCP-2)	Chose Staff (CC-1)	State IDD service systems where it is common practice to develop a service plan based on the individual's preferences is likely to also provide staffing options based on their preferences.	These two measures are positively and significantly correlated. $r = 0.344$ $p = 0.023$

Measure Being Tested	Criterion	Relational Hypothesis	Test Results
ADL Goal (PCP-3)	Community Job Goal (PCP-1)	In the support delivery system, Activities of Daily Living is a foundational element of assessment of functional support need, which is used to establish eligibility for service. Deficits in the ability to perform ADLs are therefore considered "low hanging fruit" in terms of developing a support plan. In many cases, ADL Goals may be carried over year over year for an adult receiving services, and it is unclear if they have chosen that goal. We would expect that service systems that use a deficit-based assessment and service planning may be more likely to have ADL goals in service plans. Person-centered-plans that are more progressive seek to support adults with IDD in their community employment and community participation goals, regardless of ADL deficits. We would expect to see community employment goals associated with a person's desire for community employment in progressive state service systems. We hypothesize a negative correlation between PCP-3 (ADL goal) and PCP-1 (Community job goal)	These two measures are negatively and significantly correlated. $r = -0.342$ $p = 0.024$
Lifelong Learning (PCP-4)	Has Friends (CI-2)	People with wider social circles are more likely to get exposed to new ideas and concepts. Therefore, states where people with IDD are more likely to report having friends (outside family and staff) would be expected to also have a high proportion of people with IDD reporting opportunities for lifelong learning.	These two measures are positively and significantly correlated. $r = 0.764$ $p < 0.001$

Measure Being Tested	Criterion	Relational Hypothesis	Test Results
Lifelong Learning (PCP-4)	Satisfaction with Community Inclusion Scale (PCP-5)	Exposure to new ideas and concepts would be expected to increase one's expectations of inclusion in a broader range of community activities, thus increasing the sense of "relative deprivation" and increasing dissatisfaction with one's current level of community inclusion. We would therefore expect a negative association between PCP-5 and PCP-4	These two measures are negatively and significantly correlated. r = -0.498 p = 0.002
Social Connectedness (CI-1)	Respect for Personal Space (HLR-1)	People whose personal space is respected by the people with whom they interact are more likely to feel socially connected to them. We would therefore expect a positive association between social connectedness and respect for personal space	These two measures are positively and significantly correlated. r = 0.387 p = 0.012
Transportation Availability Scale (CI-3)	Satisfaction with Community Inclusion Scale (PCP-5)	People who have readily available means of transportation are more likely to be satisfied with their level of engagement in activities outside the home. CI-3 and PCP-5 would therefore be expected to be positively associated.	These two measures are positively and significantly correlated. r = 0.404 p = 0.009
Community Inclusion Scale (CI-4)	Economic resources of the state [Source: [Source: U.S. Census Bureau]	Resource-rich states would be expected to have greater ability to support people with IDD to engage in activities outside their home. Two measures of state-level resource availability were used to test this hypothesis: per-capita income and per-capita number of jobs.	Both measures of state-level resources are positively and significantly correlated with CI-4. Per-capita income: r = 0.345 p = 0.023 Per-capita number of jobs: r = 0.471 p = 0.003
Can Change Case Manager (CC-2)	Life Decisions Scale (CC-4)	A state system that allows its clients a high degree of choice in life decisions is expected to also allow them to choose or to change their case managers, given that these two areas of choice reflect a common service philosophy.	These two measures are positively and significantly correlated. r = 0.349 p = 0.022

Measure Being Tested	Criterion	Relational Hypothesis	Test Results
Can Stay Home When Others Leave (CC-3)	Life Decisions Scale (CC-4)	A state system that allows its clients a high degree of choice in life decisions is expected to also provide them with the option of staying home alone when others leave, given that these two areas of choice reflect a common service philosophy.	These two measures are positively and significantly correlated. r = 0.552 p = 0.001
Satisfaction with Community Inclusion Scale (PCP-5)	Availability of assisted living resources in the state [Source: AARP Long-Term Services and Supports State Scorecard]	Individuals with IDD are often dependent on assistance for engaging in community activities. Therefore, states that rank higher in per capita assisted living and residential care units are expected to also rank higher on their PCP-5 score	These two measures are positively and significantly correlated. ρ=0.417 p=0.016

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Multiple studies have provided evidence that NCI data elements correlate with each other as would be logically expected and align with other measurement systems assessing similar dimensions. Statistically significant relationships in the expected direction among the state-level scores is further empirical evidence of criterion validity.

____

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

We examined the number and percentage of respondents who were excluded from the measure calculation due to exclusion criteria, as well as the percentile distributions (Min, 25th, 50th, 75th percentile, and max) across measured entities.

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance* *measure scores*)

Here is the distribution of exclusions among states:

Exclusion criterion	N excluded	% excluded	Distribution across states N=22,009 (Min, 25th, 50th, 75th percentile, and max)
For Section I items: (a) Based on survey protocol, the surveyor found that the respondent did not give any valid responses to any Section I questions	5,053	22.9%	(3%, 13%, 23%, 26%, 62%)
(b) All questions in Section I were left blank, or marked "not applicable" or "don't know".	1,882	8.6%	(0%, 0%, 5%, 14%, 42%)
For Section II items: (c) Section II was completed without using proxy, while Section I was deemed invalid (see criteria a above), or	59	0.3%	(0%, 0%, 0%, 0%, 5%)
(d) All questions in Section II were left blank, or marked "not applicable" or "don't know".	311	1.4%	(0%, 0%, 1%, 1%, 7%)

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis.* ***Note****:* ***If patient preference is an exclusion****, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

- Exclusion (a), (the surveyor found that the respondent did not give any valid responses to any Section I questions), accounting for 22.9% of all surveys, represents the majority (69%) of all exclusions and is meant to safeguard the validity of measures that utilize Section I items. This exclusion, with its conservative approach, prevents the inclusion of responses with sub-standard reliability in measure calculations.
- Exclusion (b) (All questions in Section I were left blank or marked "not applicable" or "don't know") accounts for about 8.6% of all surveys and represents about a quarter (26%) of all exclusions. It is purely objective and is needed to prevent the inclusion of responses that do not contribute meaningful data for Section I items.
- Exclusion (c) (Section II was completed without using proxy, while Section I was deemed invalid) only accounts for 0.3% of all surveys and represents less than 1% of all exclusions. However, it is in place to safeguard the validity and reliability of measures that utilize Section II items by excluding responses provided by an individual whose responses to Section I were assessed as unreliable. Given that a very small percentage of individuals are being excluded, it is unlikely this exclusion unduly affects the measure score.
- Exclusion (d) (All questions in Section II were left blank or marked "not applicable" or "don't know") accounts for 1.4% of all surveys, 4% of all exclusions. Its determination is purely objective and is needed to

prevent the inclusion of surveys that do not contribute meaningful data for Section I items.

- Exclusions are based on the uniformly applied criteria, most of which are objective and all of which are standardized. The exclusions were put in place to ensure accurate calculation of the measures and to safeguard validity and reliability. It is important to note that to the extent possible, exclusions eliminate unreliable responses, not the entire survey. For example, a survey where Section I responses are excluded from measure calculations may still be included in measures based on Section II items if Section II responses were provided by a proxy.

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b3.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 13 risk factors**

☒ **Stratification by 5 (residence type) risk categories**

☐ **Other,**

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

The final model was used to adjust both the *Life Decisions Scale* and the *Community Inclusion Scale*. The regression results for the two adjusted measures are provided in the **2b3.1 tab of the Supplemental Tables attachment**. The tabulated coefficients and model parameters were obtained through OLS regression analysis.

Method of Risk Adjustment for *Life Decisions Scale* and *Community Inclusion Scale*

The same risk-adjustment model is used for both scales. The model includes 13 predictors. With the exception of age, a continuous variable measured in years, all predictors are dummy-variables coded from categorical factors. All dummy variables are coded 0 if the category is absent and 1 if the category is present for the individual participant. The individual-level model equation for both scales is:

Predicted scale value=

$$\beta_0 + (\beta_1 * \text{age}) + (\beta_2 * \text{mild ID}) + (\beta_3 * \text{moderate ID}) + (\beta_4 * \text{severe ID}) + (\beta_5 * \text{profound ID}) +$$
$$(\beta_6 * \text{unspecified ID level}) + (\beta_7 * \text{non-spoken communication}) + (\beta_8 * \text{some behavioral supports}) +$$
$$(\beta_9 * \text{extensive behavioral supports}) + (\beta_{10} * \text{non-ambulatory}) + (\beta_{11} * \text{ambulatory with aids}) +$$
$$(\beta_{12} * \text{excellent or very good health}) + (\beta_{13} * \text{proxy used})$$

A state's adjusted score for the measure is calculated as follows:

(state's mean observed score) – (state's mean predicted score) + (observed grand mean score)

where the observed grand mean score is the mean of the measure over the entire individual-level dataset (data pooled from all member states).

2b3.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

The following measures are not risk-adjusted:

- Community Job Goal
- Person-Centered Goals
- ADL Goal
- Lifelong Learning

- Satisfaction with Community Inclusion Scale
- Social Connectedness
- Has Friends
- Transportation Availability Scale
- Chose Staff
- Change Case Manager
- Can Stay Home When Others Leave
- Respect for Personal Space Scale

For all of the above measures, the justification for non-adjustment is conceptual. In all cases, the individual-level measure scores were associated with most or all of the candidate factors considered for risk adjustment. However, within the context of HCBS measures, and especially given our emphasis on assessing *person-centered* practices, the association with individual characteristics should be considered a reflection of the quality of services. For example, we found that *Lifelong Learning* is negatively associated with age and level of intellectual disability, and positively associated with overall health status. However, these associations are indicators that services and supports are not sufficiently tailored to the personal characteristics of the individual; that is, they are not sufficiently person-centered. In these cases, adjusting for individual characteristics would have limited the ability of the measure to fully reflect the quality of the service, given that *service quality*, in this context, incorporates person-centeredness.

One possible exception to the above justification was the *Transportation Availability Scale*. For this measure, we considered the rurality of the community to be spurious to service quality and a candidate for adjustment. However, statistical analysis (analysis of variance and t-tests for differences between means) found no significant association between the scale score and the rurality of the zip code area. We therefore decided to leave that measure unadjusted as well.

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

A list of candidate risk adjusters were identified based on a combination of literature review and exploratory statistical analysis (Pearson correlations, Chi-squared tests, t-tests for differences between means, z-tests for differences between proportions) that suggested that they were associated with the outcome(s) being measured. Statistical significance at $p<0.05$ was the threshold for considering a factor as a candidate for risk adjustment. Starting with the full list of candidate factors, the NCI team arrived at a parsimonious individual-level model by eliminating redundancies while retaining optimal model fit, after estimating and examining the results of multiple OLS regression models. The dependent variables for these models was the measure being adjusted and the predictors were the potential adjustment factors. We then tested between-state variability of the predictors in the final individual-level model in order to eliminate from the list of adjusters those that did not vary significantly across states.

Sociodemographic and clinical characteristics were treated similarly in identifying candidate adjustment factors and subsequent analyses to select the final adjustment methodology.

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

☒ **Published literature**

☒ **Internal data analysis**

☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

See **2b3.1 tab of the Supplemental Tables attachment** for the statistical results of the final models used to risk-adjust the *Life Decisions* and *Community Inclusion* scales.

Distributions of the unadjusted and adjusted scales are displayed in Exhibits 4 – 7 below.

Exhibit 4. Distribution of the Unadjusted State-Level *Life Decisions Scale*

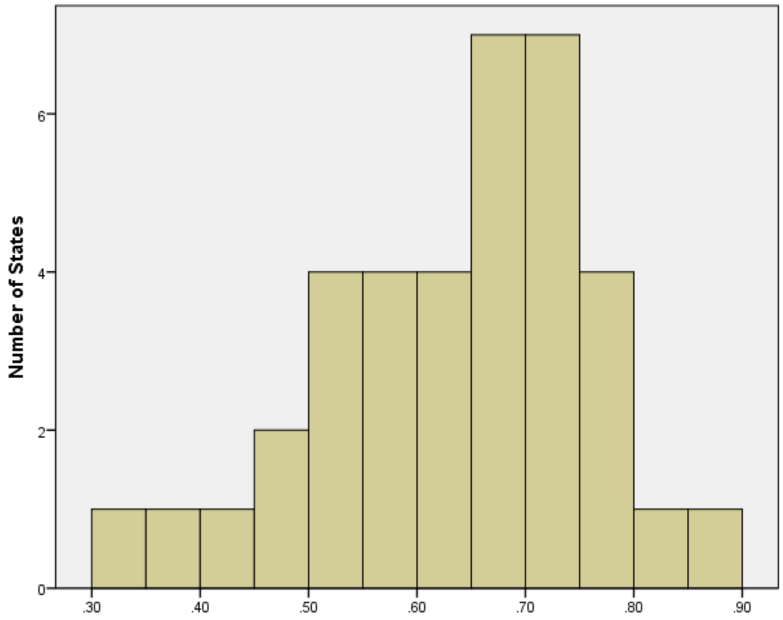


Exhibit 5. Distribution of the Adjusted State-Level *Life Decisions Scale*

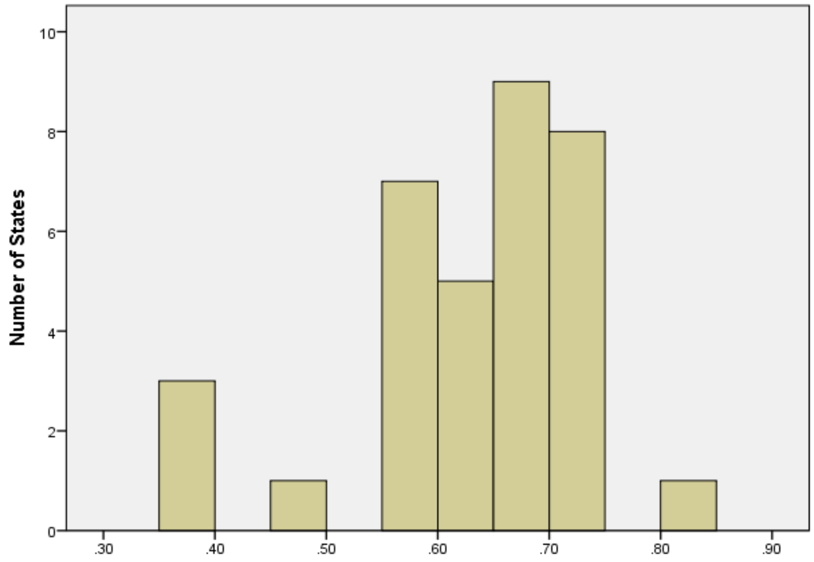


Exhibit 6. Distribution of the Unadjusted State-Level *Community Inclusion Scale*

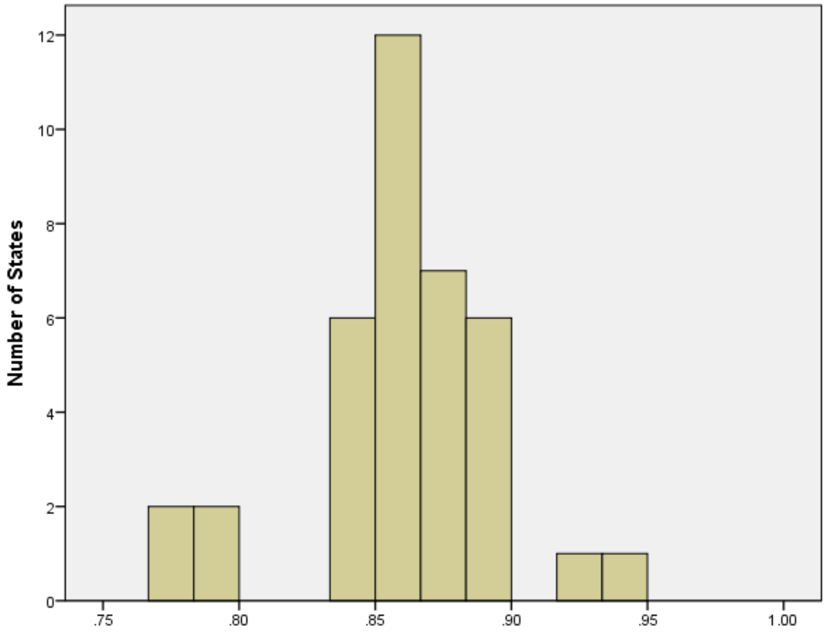
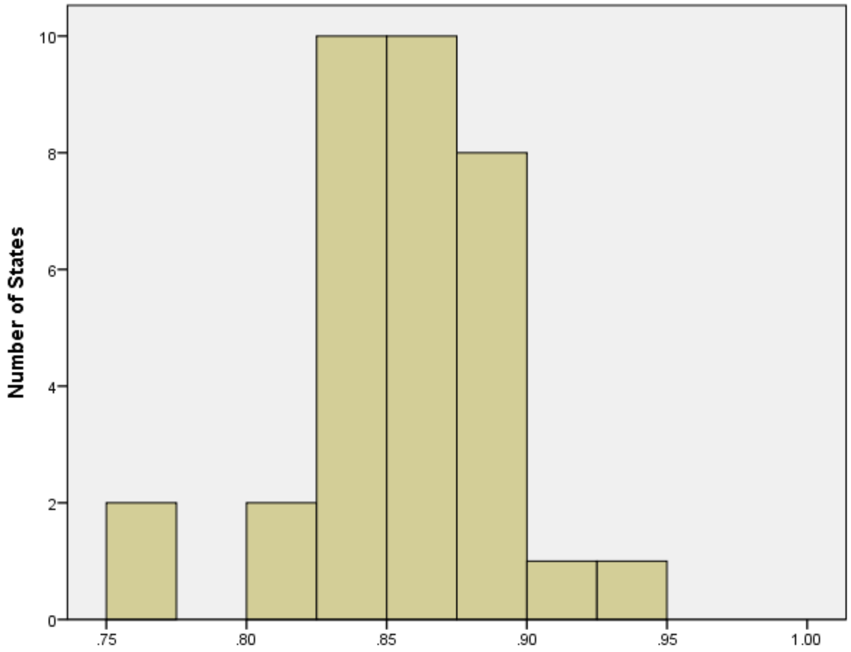


Exhibit 7. Distribution of the Adjusted State-Level *Community Inclusion Scale*



2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

Social risk factors were identified and final selections made through the same process used for patient risk factors as described above.

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Our risk adjustment models demonstrate reasonable predictive validity for *Life Decisions Scale*(#CC-4) and *Community Inclusion Scale*(#CI-4). Using multiple linear regressions, we conducted regression diagnostic analyses to assess model performance, examining predictive ability, and outlier influence.

If stratified, skip to 2b3.9

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

The R-Squared values were 0.159 for *Life Decisions Scale*(#CC-4) and 0.085 for *Community Inclusion Scale*(#CI-4).

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

We conducted Outlier Influence Analysis to identify any outlier observations with potential large effects on *Life Decisions Scale*(#CC-4) and *Community Inclusion Scale*(#CI-4) scores. Specifically, we calculated Cook's distances (D scores), where a value of 1.0 or higher would suggest an influential observation. We found that all Cook's D scores were under the threshold of 1.0. The maximum D scores were 0.0011 for *Life Decisions Scale*(#CC-4) and 0.0020 for *Community Inclusion Scale* (#CI-4), respectively.

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

We split our sample into deciles and for each decile, we calculated the mean expected and observed values for the two risk-adjusted measures, *Life Decisions Scale* (CC-4) and *Community Inclusion Scale* (CI-4). We then calculated the ratio of predicted to observed mean values for each decile. A ratio of 1 would indicate perfect agreement between the observed and predicted values, indicating perfect predictive performance. Details of these calculations for the two risk-adjusted measures are provided in **tab "2b3.8" of the Supplemental Tables attachment**. Overall, there is fairly good agreement between predicted and observed mean values, indicating acceptable model performance.

2b3.9. Results of Risk Stratification Analysis:

Risk-adjusted *Life Decisions* and *Community Inclusion Scales*, are further stratified by 5 residential setting categories:

- Own home or apartment
- Parents' or relatives' home
- Foster care or host home
- Group residential setting (e.g., group home)
- Intermediate Care Facility for Individuals with Intellectual Disabilities (ICF/IID), nursing facility, or other institutional setting

There are both conceptual/policy and empirical reasons for this stratification. Conceptually, the need for types and mixes of HCBS supports vary by residential setting, impacting the interpretation and program/policy implications of outcomes. Providing scores for each residential setting separately provides states with meaningful information about the outcomes of these different service/support strategies, offering detailed, actionable recommendations for improvement. Further, risk-adjusted measures significantly vary by residential setting, providing empirical support for the informational value of reporting these measures separately for the 5 settings. Exhibit 8 below shows the mean measure scores by residential setting and summary ANOVA results for the two stratified measures.

Exhibit 8. Mean Scores of *Life Decisions* and *Community Inclusion Scales* by Residential Setting

Residence	Life Decisions Scale	Community Inclusion Scale
Own home or apartment	0.8030	0.8672
Parents' or relatives' home	0.7004	0.8561
Foster care or host home	0.6114	0.8749

Residence	Life Decisions Scale	Community Inclusion Scale
Group residential setting	0.5146	0.8558
Institutional setting	0.3026	0.7644
Eta-squared	0.153	0.011
p-value	<0.001	<0.001

A state's risk-adjusted scores stratified by residential setting are calculated as follows:

1. Calculate the mean predicted value of the measure separately for each residential setting
2. Calculate the mean observed value of the measure separately for each residential setting
3. Calculate the observed grand mean of the measure over the full dataset (pooled data from all states)
4. Calculate the risk-adjusted score for each residential setting within the state as:
   (the residential setting's mean observed score) – (the residential setting's mean predicted score) + (observed grand mean score)

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

Overall, results of the statistical analyses suggest that our risk-adjustment model has sufficient ability to predict the two measures that are being adjusted. We interpret the results as indicating that these two measures are reasonably well adjusted for sociodemographic and clinical characteristics of program participants.

2b3.11. Optional Additional Testing for Risk Adjustment (***not required****, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

The grand mean of each measure calculated over the full dataset (pooled data from all states) is used as the benchmark for identifying meaningful differences in states' performance on that measure. A 95% confidence interval is constructed around the grand mean and around each state's mean. States whose confidence interval is below that of the grand mean are considered to have lower performance than the benchmark. States whose confidence interval is higher than that of the grand mean are considered to have higher performance than the benchmark. States whose confidence interval overlaps with the confidence interval of the grand mean are considered to be performing at the benchmark.

Provided that member states use similar sampling designs from one year to the next, meaningful change in a measure from year to year can be identified by determining whether the confidence interval of a measure in a given year is below, above, or overlapping with the confidence interval of the measure from a past year.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Numbers and percentages of states performing below, above, and at the benchmark are tabulated in **tab 2b4.1 of the Supplemental Tables attachment**.

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

The table in in **tab 2b4.1 of the Supplemental Tables attachment** indicates that the benchmarking method described above is able to identify statistically significant and meaningful performance differences among states. For the 14 measures being proposed, between 8% and 30% of the states performed below the benchmark, between 14% and 60% of the states performed above the benchmark, and between 19% and 76% of the states performed at the benchmark.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note*: This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i*.e., what do the results mean and what are the norms for the test conducted*)

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

In 2020, the NCI team, with participation from two member states, conducted a **non-responder study** comparing the characteristics of respondents to those of sampled individuals who did not respond to the survey in the 2018-2019 survey cycle. Background information for non-responders was provided by the states based on administrative records. This analysis assessed "attrition bias" which can be introduced into data collected from a representative sample if everyone in the sample cannot be interviewed and those who could not be interviewed share characteristics relevant to the analysis. For example, if males in the sample are less likely to participate in the survey, then the data underrepresents males even though the original sample did not. If responders and non-responders have similar characteristics, then attrition bias is low or non-existent and the data are representative of the target population with respect to relevant characteristics.

The study included 429 respondent and 363 non-respondent cases from one state and 2,137 respondent and 1,702 non-respondent cases from the second state. The following factors were compared between responders and non-responders:

- demographic characteristics
- type & extent of disability
- residence type
- legal status (legal competence vs. guardianship)
- funding source
- participation in self-directed supports & services.

Differences between the two groups were tested for statistical significance using z-tests for difference between proportions or Chi-squared tests in the case of categorical variables and two-sample t-tests for difference between means in the case of continuous variables.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse;* ***if no empirical sensitivity analysis****, identify the approaches for handling missing data that were considered and pros and cons of each*)

State 1: Of the 792 sampled individuals, 54.2% responded and 45.2% did not respond. Of the 247 non-responses for which a reason was given, 57% could not be reached for a variety of reasons, 41% were refusals by a caregiver, and 2% were refusals by the person themselves.

State 2: Of the 3,839 sampled individuals, 55.7% responded and 44.3% did not respond. Of the 1,702 non-responses, 60% could not be reached, 26% were refusals by people other than the person, 9% were refusals by the person, and 9% were cancellations or no-shows after initially agreeing to respond.

Comparison results in State 1

Responders were significantly…

- Older, on average, than non-responders (44 vs. 35)
- More likely to have mild or moderate (as opposed to severe or profound) intellectual disability (70% vs. 55%)
- More likely to prefer spoken (as opposed to non-spoken) means of communication (80% vs. 74%)
- More likely to live in a group setting (36% vs. 6%) or a foster home (15% vs. 3%) and less likely to live in parents'/relatives' home (35% vs. 77%)
- More likely to have a legal guardian (76% vs. 71%)
- More likely to receive supports for a behavioral problem (53% vs. 42%)
- Less likely to use self-directed supports option (22% vs. 63%)

No significant differences were found in

- Gender
- Race
- Funding source

Comparison results in State 2

Responders were significantly…

- Older, on average, than non-responders (44 vs. 39)
- More likely to live in a group setting (54% vs. 39%) and less likely to live in parents'/relatives' home (25% vs. 36%)
- More likely to have Medicaid HCBS funding (92% vs. 84%)
- Less likely to use self-directed supports option (6% vs. 11%)

No significant differences were found in

- Gender

- Race
- Level of intellectual disability
- Preferred means of communication (spoken vs. non-spoken)
- Legal status (legal competence vs. guardianship)

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias**?** (i*.e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted;* ***if no empirical analysis****, provide rationale for the selected approach for missing data*)

The two states where the non-responder study was conducted had the following shared results:

- The leading cause for non-response was inability to contact the sampled individual.
- The data available for analysis overrepresented older people and people who live in group settings while underrepresenting people who live in parents'/relatives' homes and people who receive self-directed supports and services.
- There were no significant differences between responders and non-responders in gender and race.

All other comparisons had differing results in the two study states.

The shared results suggest recommendations to states in maintaining up-to-date contact information for their service users and making concerted efforts to reach younger and more independent (e.g., living in parents'/relatives' homes or self-directing) consumers selected into their samples. Both of the study states have rural counties and our communications with state officials suggest that at least some of the failure to contact individuals in rural areas is due to the difficulty in arranging transportation to visit their homes. Conducting the IPS via videoconferencing may, to some degree, address this challenge and improve response rates. The NCI team has recently completed a pilot study to test the reliability and validity of IPS data collected via videoconference rather than direct-contact. Preliminary results are encouraging, suggesting that this will be a feasible alternative in hard-to-reach areas, improving their response rates.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. Data Elements Generated as Byproduct of Care Processes.

Other

If other: In-person survey

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. To what extent are the specified data elements available electronically in defined fields (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.

Patient/family reported information (may be electronic or paper)

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

As PRO measures, the data are collected directly from the person enrolled in HCBS services and/or their proxy. Because the measures are experience of care measures, they are not located in electronic sources.

3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.

Attachment:

3c. Data Collection Strategy

> Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

At the present time, all NCI for ID/DD HCBS Measures are collected by 38 states every year, on average, thus demonstrating the challenges to feasibility can be overcome. During the first year of data collection, states reported challenges setting up the initial contact system, establishing the sample that will meet the state’s quality management needs, and coordination of scheduling the interviews. Most states report this as an initial challenge which would be overcome once the initial process and protocols are established, and through repeated years of data collection. States under contract with external administrative entities (local county administrative agencies, managed care plans) report challenges with establishing agreements to release specific individual data due to confidentiality concerns. NCI surveyors must be trained to the NCI protocol each year, however most states experience returning surveyors, and as such, the annual NCI training is a refresher course requiring less time than the initial course. NCI is recognized as part of the Medicaid Quality Management system, and therefore the state’s access to data is addressed in Medicaid consent regulations. States typically address the issue of data access at the time of contract renewal with the administrative entity. Sample identification provides a challenge for those states electing to oversample and/or stratify by specific population (region within the state for example) and to assure adequate coverage of the variety of service types available within the state Developmental Disability HCBS system. However, NCI national team provides technical assistance in the sample identification procedures, and once the initial sample pool is identified, states collecting data annually utilize the same sample identification procedure for subsequent years. Individuals who receive Home and Community based Long-Term services and supports (HCBS-LTSS) who provide the data are asked to voluntarily participate in the data collection; the data collection interview requires less than an hour to complete. The instrument is adapted to refer to names of people and places (case manager name for example) to increase ease of understanding for the enrolled individuals participating. States collecting the information must coordinate the scheduling of the data collection time and location; most states work closely with case management agencies to facilitate the ease of this process.

3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*.

The NCI for ID/DD HCBS Measures data collection at the state level occurs in the context of an agreement between states that participate as members of the National Core Indicators through the National Association of State Directors of Developmental Disability Services (NASDDDS). An annual membership fee of $14986 is required to participate in data collection. Membership fees provide participating states access to the technical assistance supporting an annual work plan. The National Core Indicators team ensures fidelity of implementation of the survey through assistance with details of sample identification, training for surveyors and shadow surveying for quality monitoring, guidance on assent/consent processes. Additionally, the membership fee allows access to a secure Online Data ENTRY SYSTEM (ODESA), data cleaning, analysis, risk adjustment where appropriate, development of reports, posting and hosting of publicly available reports on the national core indicators website, as well as a chart generator.

De-identified data are available to researchers for secondary data analyses. The costs for data access are determined by the nature and scope of the data request and the technical assistance costs that are anticipated by the NCI team.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
Quality Improvement (external benchmarking to organizations)	Public Reporting Medicaid Adult Core Health Care Quality Measure Set https://www.medicaid.gov/medicaid/quality-of-care/performance-measurement/adult-and-child-health-care-quality-measures/adult-health-care-quality-measures/index.html Regulatory and Accreditation Programs Connecticut Medicaid 1915(c) HCBS Waiver Assurances https://portal.ct.gov/-/media/DDS/events2020/EDS_Waiver_Renewal_10_27_2020.pdf Indiana Family and Social Services Administration Medicaid 1915(c) HCBS Waiver Assurances http://ai.org/fssa/files/CIH%20Final-Approved.pdf Quality Improvement (Internal to the specific organization) Arizona: Monitored provider rates as the state incentivized Community and Supported Employment initiatives https://www.aaidd.org/publications/bookstore-home/product-listing/cross-cultural-quality-of-life-enhancing-the-lives-of-people-with-intellectual-disability-second-edition Massachusetts: The state Department of Developmental Services used measure of loneliness and friendships to identify areas for quality improvement in supporting relationships https://shriver.umassmed.edu/wp-content/uploads/2020/07/QINA-Friendship_final_web2.pdf Kentucky Division of Developmental and Intellectual Disabilities reviews outcomes from National Core Indicators with stakeholders to identify recommendations for quality improvement http://www.kentuckycq.org/wp-content/uploads/2018/02/QI-draft-recom_2017_final.pdf

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

d. Public Reporting

- Names of program and sponsor:

Arkansas Developmental Disability Services, Colorado DHCPF/OCL, Florida Agency for Persons with Disabilities, Georgia DBHDD, Illinois DDD, Indiana Disability and Rehab Services, Kansas Aging and Disability Services, Kentucky DD Services, Maine Office of Aging and Disability Services, Massachusetts Dept. of Developmental Services, Minnesota Disability Services Division, Missouri Division of Developmental Disabilities, Nebraska Division of Developmental Disabilities, Nevada Aging and Disability Services, New Jersey Division of Developmental Services, North Carolina MH/DD/SAS, Ohio Dept. of Developmental Disabilities, Oklahoma Developmental Disabilities Services, Oregon Developmental Disability Services, Pennsylvania Office of

Developmental Programs, Rhode Island Division of Developmental Disabilities, Tennessee Department of intellectual and Developmental disabilities, Utah Service for People with Disabilities, Virginia DBHDDS, Wisconsin Division of Long-Term Care /DHS, and Wyoming Developmental Disabilities Section DHF

- Purpose: Reporting in Medicaid Adult Core Set 2019
- Geographic area and number and percentage of accountable entities and patients included: Twenty-six states across the country (100% of accountable entities of the survey year). A total of 6,828 patients were included (99.6% of all respondents with valid surveys).
- Level of measurement and setting: As mentioned, the NCI for ID/DD HCBS Measures are relevant in the state HCBS setting. The measure scores are reported at the state-level.

f. Quality Improvement (Internal to the specific organization)

- Names of program and sponsor: Arizona Department of Economic Security – Division of Developmental Disabilities.
- Purpose: Monitoring of employment goals to incentivize individualized employment supports.
- Geographic area and number and percentage of accountable entities and patients included: Sample of individuals who are receiving supports from HCBS supports in Arizona.
- Level of measurement and setting: State level monitoring
- Names of program and sponsor: Massachusetts Department of Developmental Disabilities Services
- Purpose: Monitoring of relationship outcomes (loneliness, friends who are not staff) as a multi-year quality improvement effort is underway.
- Geographic area and number and percentage of accountable entities and patients included: Sample of individuals who are receiving supports from HCBS supports in Massachusetts (statewide). 400 individuals typically included in sample, representing approximately 24,000 adults receiving HCBS supports
- Level of measurement and setting: State level monitoring
- Names of program and sponsor: Kentucky Division of Developmental and Intellectual Disabilities
- Purpose: Review of outcomes for individuals with a stakeholder group and identification of quality improvement recommendations. Emphasis on person-centered thinking, planning, and supports.
- Geographic area and number and percentage of accountable entities and patients included: Sample of individuals who are receiving supports from HCBS supports in Kentucky (statewide) 400 individuals typically included in sample, representing approximately 18,000 adults receiving HCBS supports.
- Level of measurement and setting: State level monitoring

g. Regulatory and Accreditation Programs

- Names of program and sponsor: Connecticut Department of Developmental Services, Employment and Day Supports Waiver
- Purpose: Demonstrate compliance with Statutory Assurances required through Federal Medicaid Regulations found at 441.302 and Performance Measures required in the HCBS 1915(c) Application and Technical Guide.
- Geographic area and number and percentage of accountable entities and patients included:
- Level of measurement and setting: Statewide

Names of program and sponsor: Indiana Family and Social Services Department Community Integration and habilitation Waiver

- Purpose: Demonstrate compliance with Statutory Assurances required through Federal Medicaid Regulations found at 441.302 and Performance Measures required in the HCBS 1915(c) Application and Technical Guide.
- Geographic area and number and percentage of accountable entities and patients included:
- Level of measurement and setting: Statewide

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

n/a

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

In addition to all aforementioned cases of public reporting and use, NCI survey results are also publicly available on the NCI website(https://nationalcoreindicators.org/survey-reports).

4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

As measured entities, states receive performance results in the form of annual reports. All participating states (46 states and Washington D.C.) are included-- there is no sampling process.

Data and codebooks are provided to all participating states in the form of comma delimited files and pdf files, respectively. All participating states are included.

Assistance with interpretation is offered to all participating states as part of membership in NCI; Custom reports, analysis, and presentations are available to states by request.

In addition, to collect feedback directly from survey respondents, a cognitive interview using the survey instrument was conducted in 2019. The process and findings are described below.

4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

Performance results are provided in the form of annual reports, and provided to all participating states as well as the general public via accessible, 508-compliant pdf reports on the NCI website.

As mentioned, data and codebooks are provided to all participating states in the form of comma delimited files and pdf files, respectively. These files are transmitted over encrypted, HIPAA-compliant methods.

Assistance with interpretation is offered to all states as part of membership in NCI; Custom reports, analysis, and presentations are available to states by request.

Each year, a NCI Annual Meeting is held for participating states and survey staff, where educational/explanatory presentations and workshops are provided free of charge, and feedback and input is sought on any surveying changes being considered.

A cognitive interview was conducted in 2019. The cognitive interview survey instrument contained the critical elements of the 14 measures as well as some questions lead-in and background questions. The surveyor completed a brief “questionnaire” after each response, recording whether or not the respondent asked the question to be repeated and/or clarified, the nature of the clarification, and whether or not the response options were clear (and if not clear, the surveyor suggested revisions). After each critical-element question, the respondent was asked how difficult or easy it was to provide a response. If the question was characterized as difficult, an open-ended follow-up question was asked about the source of the difficulty. The respondent feedback questions also included open-ended items to probe how the key concept(s) contained in the question were interpreted.

The cognitive interview protocol was reviewed and approved by HSRI’s Institutional Review Board. Three states participated by recruiting volunteers from among their service recipients, the typical respondent base for the NCI Adult In-Person Survey. Participants received a stipend of $40. The interviews were conducted by members of the measure development team. Ten volunteers participated in the cognitive interview. Six participants lived in a group home or other congregate setting and four lived in their own or family home.

A mixed-methods approach was used to analyze the results of the cognitive interview. Surveyor and participant feedback questions with forced-choice responses were subjected to quantitative analysis. Frequency distributions and descriptive statistics were used to analyze and interpret the results. The “framework method” (Gale et al., 2013) was used for the qualitative analysis of open-ended feedback questions from the surveyors and participants. Recurring themes were coded into a thematic framework and formed into a matrix for analysis and interpretation.

Gale, N.K., Heath, G., Cameron, E., Rashid S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Medical Research Methodology 13, 117. https://doi.org/10.1186/1471-2288-13-117

4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

Feedback from states (measured entities):

The NCI team regularly obtains feedback from the states via multiple venues:

- HSRI surveys state members to determine satisfaction with the measures and the process overall.
- As measures were developed in new areas, including measures of person centered practices, the team held focus groups with state service coordinators and managers in the developmental disabilities systems to learn about measurement gaps and needs that exist, and how measures of person centered practice are performed.
- The National Association of State Directors of Developmental Disability Services, (NASDDDS) research committee serves as an advisory body and oversees the direction and use of NCI, including providing feedback from states on priorities for use of measures.
- Participating states submit annual workplans that include feedback on the entire data collection and reporting process.
- Feedback is provided by state officials and surveyors at NCI trainings
- The survey instrument includes questions to surveyors about the overall survey experience and solicits their recommendations for improving the experience.
- Feedback from states about their overall experience with the measures and with the data collection process are obtained during annual NCI Conferences

Feedback from Survey Respondents:

A series of cognitive interviews were conducted in 2019. The cognitive interview survey instrument contained the critical elements of the 14 measures as well as some questions lead-in and background questions. The surveyors completed a brief “questionnaire” after each response, recording whether or not the respondent asked the question to be repeated and/or clarified, the nature of the clarification, and whether or not the response options were clear (and if not clear, the surveyors suggested revisions). After each critical-element question, the respondents were asked how difficult or easy it was to provide a response. If the question was characterized as difficult, an open-ended follow-up question was asked about the source of the difficulty. The respondent feedback questions also included open-ended items to probe how the key concept(s) contained in the question were interpreted.

The cognitive interview protocol was reviewed and approved by HSRI’s Institutional Review Board. Three states participated by recruiting volunteers from among their service recipients, the typical respondent base for the NCI Adult In-Person Survey. Participants received a stipend of $40. The interviews were conducted by members of the measure development team. Ten volunteers participated in the cognitive interview. Six participants lived in a group home or other congregate setting and four lived in their own or family home.

A mixed-methods approach was used to analyze the results of the cognitive interview. Surveyor and participant feedback questions with forced-choice responses were subjected to quantitative analysis. Frequency distributions and descriptive statistics were used to analyze and interpret the results. The "framework method" (Gale et al., 2013) was used for the qualitative analysis of open-ended feedback questions from the surveyors and participants. Recurring themes were coded into a thematic framework and formed into a matrix for analysis and interpretation.

Gale, N.K., Heath, G., Cameron, E., Rashid S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Medical Research Methodology 13, 117. https://doi.org/10.1186/1471-2288-13-117.

4a2.2.2. Summarize the feedback obtained from those being measured.

Feedback from states (measured entities)

In response to the 2020 state survey, 22 states provided written feedback to NCI. Nineteen out of 22 states responded they were satisfied or very satisfied, and 3 states indicated a neutral response. When asked if NCI is a good value for the state's DD system, 13 states strongly agreed, 8 states agreed, and 1(one) state disagreed.

State surveys and state meetings at the Annual NCI Conference provide the NCI team with recommendations for new domains and measures to be added, survey questions that are of limited value and can be removed, and revisions to existing survey items (both question wording and response options).

Surveyor feedback section of the Adult In-Person Survey and the regular surveyor training sessions provide feedback on question wording as well as instructions to surveyors included in the survey instrument and the manual.

4a2.2.3. Summarize the feedback obtained from other users

Feedback from Survey Respondents: Cognitive Interview Results

A third of the questions had a request for repetition and slightly over half of the questions had a clarification request from at least one of the 10 participants. About 47% of the questions had no repetition or clarification request from any of the respondents. Most of the clarification requests were for additional information that surveyors are routinely trained to provide, such as alternative wording or examples for the key concept(s). Question difficulty was rated on a scale of 0 (very easy) to 3 (very difficult). The mean score over all questions and all participants was 1.1 and none of the questions had a mean score higher than 1.5. These results indicate that overall, the participants found the questions to be easy to understand and to answer. Qualitative analysis of open-ended follow-up questions about why the question was "difficult" revealed that for the majority of participants, the source of the difficulty was not cognitive burden. Commonly offered reasons were emotional (e.g. the question on loneliness triggered sadness and made it difficult to answer) or situational (e.g. respondent was tired).

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

Feedback from States:

Feedback from the states on all aspects of the data collection, analysis, and reporting process is systematically recorded and used in revising the survey. These include state recommendations for new items to add and measures to revise/update. This is documented and retained by the National Core Indicators team for consideration on an annual basis. The instrument may be modified for clarity on an annual basis and the team uses feedback to inform training for surveyors.

Measures are generally retained for a 4-year period to allow for consistency and for tracking of performance over time. In the four year review phase, the NCI team reviews all domains in detail, examining the performance of items.

Feedback from Survey Respondents:

Overall, the cognitive interview did not reveal any major difficulties in comprehension or wide variation in how the

questions were interpreted. The results, especially from the qualitative analysis, suggest a few minor revisions to question wording or the response options which will be included in planning discussions for the next revision cycle.

Improvement
Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

The measure #PCP-1 is used by the Supported Employment Leadership Network (SELN) to measure each participating state's progress in utilizing supported employment services. Missouri DD agency, for example, utilized this measure in combination with other NCI measures of those who report they want a job, to improve rates of employment across the state. Missouri rates of people indicating they had a paid community job increased from 9% in 2017-18 to 14% in 2018-19.

Between 2014 and 2020, 20 additional states have joined as NCI members and the total number of individual surveys entered per year has increased from 13,157 in the 2012-13 reporting cycle to 22,009 in 2018-19. NCI has continuously increased its reach and application across state DD systems.

4b2. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

There were no unintended findings identified.

4b2.2. Please explain any unexpected benefits from implementation of this measure.

There were no unintended benefits identified.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria **and** there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

Yes

5.1a. List of related or competing measures (selected from NQF-endorsed measures)

2967 : CAHPS® Home- and Community-Based Services Measures

5.1b. If related or competing measures are not NQF endorsed please indicate measure title and steward.

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):

Are the measure specifications harmonized to the extent possible?

No

5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

NQF 2967 - CAHPS Home and Community Based Services Measures could be used to survey the same population as it is described as a cross-population survey. NCI for ID/DD HCBS Measures, on the other hand, were specifically designed to survey the target population of adults with intellectual or developmental disabilities who are receiving HCBS. That said, the NCI for ID/DD HCBS Measures do not have the same focus as HCBS-CAHPS measures. One area which merits mention is the transportation item because it may appear to be related with a similar focus. The Transportation availability scale that is in this measure set includes a measure of having transportation available when needed. This is not the same measure as the "Transportation to Medical Appointments" scale that exists as part of HCBS-CAHPS, which only focuses on medical appointments. Home and Community Based Services (HCBS) are intended to support people to live a life in the community that extends beyond merely medical appointments, therefore a measure of broader access to transportation is important to have.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);

OR

Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

We do not know of any NQF-endorsed measures that conceptually address both the same measure focus and the same target population.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): Human Services Research Institute

Co.2 Point of Contact: Henan, Li, hli@hsri.org, 617-876-0426-2341

Co.3 Measure Developer if different from Measure Steward: Human Services Research Institute

Co.4 Point of Contact: Henan, Li, hli@hsri.org, 617-876-0426-2341

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

Team directly involved in Measure Development for NQF

Henan Li, PhD, Human Services Research Institute

Alexandra Bonardi, MS OT, MHA, Human Services Research Institute

Nilufer Isvan, PhD, Human Services Research Institute

Valerie J. Bradley, President Emerita, Human Services Research Institute

Julie Bershadsky, PhD, formerly at Human Services Research Institute

Dorothy Hiersteiner, MPP, Human Services Research Institute

Stephanie Giordano, PhD, Human Services Research Institute

Gary Smith, Human Services Research Institute

Sara Galantowicz, PhD, Abt & Associates – Contracted consultant supporting prioritization of measures for submission to NQF for endorsement.

National Association of State Directors of Developmental Disabilities Services (NASDDDS)

Mary P. Sowers, Executive Director

Mary Lou Bourne, Chief Quality and Innovation Officer

Mary Lee Fay, Executive Director, retired

Laura Vegas, Director of National Core Indicators

Chas Mosely, NASDDDS

Robert Gettings, NASDDDS

Additional researchers and technical experts contributing to measure development discussions:

Amy Hewitt, PhD, University of Minnesota

Renata Ticha, PhD, University of Minnesota

Brian Abery, PhD, University of Minnesota

Original survey development team

John Ashbaugh, Human Services Research Institute

Sarah Taub, Human Services Research Institute

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 1997

Ad.3 Month and Year of most recent revision: 05, 2018

Ad.4 What is your frequency for review/update of this measure? Reviewed annually for reliability and validity. Major revision every 4 years.

Ad.5 When is the next scheduled review/update for this measure? 2022

Ad.6 Copyright statement: NCI® and National Core Indicators® are registered trademarks of the NASDDDS and HSRI. The NCI measures and specifications were developed by and are owned by the National Association of State Directors of Developmental Disabilities Services (NASDDDS) and Human Services Research Institute (HSRI). NASDDDS and HSRI hold a copyright on all materials associated with the NCI measures and specifications and may rescind or alter these measures and specifications at any time. Users of the NCI measures and specifications shall not have the right to alter, enhance, or otherwise modify the NCI measures and specifications or associated materials. Anyone desiring to use or reproduce the contents of reports, inclusive of data results, without modification for a non-commercial purpose, may do so without obtaining approval from NCI. The use or reproduction of NCI survey instruments and questions requires prior approval by the NASDDDS and HSRI. All commercial uses or requests for alteration of the measures and specifications must be approved by NASDDDS/HSRI and are subject to a license at the discretion of NASDDDS/HSRI. NCI measures and specifications are not clinical or disability services guidelines, do not establish a standard of medical care, nor a standard for disability services and are not intended or tested for all potential applications.

The measures and specifications are provided “as is” without warranty of any kind. NASDDDS and HSRI make no representations, warranties, or endorsements about the suitability or utility of any product, test, or protocol identified as deriving from or based on an NCI measure or specification. NCI also makes no representations, warranties, or endorsements about the quality of any agency of a state, contractor of a state agency, or other organization who uses, applies, or reports NCI performance measures. NASDDDS/HSRI has no liability to anyone who relies on NCI measures and specifications or data reflective of performance under such measures and specifications.

Ad.7 Disclaimers: The measures and specifications are provided “as is” without warranty of any kind. NASDDDS and HSRI make no representations, warranties, or endorsements about the suitability or utility of any product, test, or protocol identified as deriving from or based on an NCI measure or specification. NCI also makes no representations, warranties, or endorsements about the quality of any agency of a state, contractor of a state agency, or other organization who uses, applies, or reports NCI performance measures. NASDDDS/HSRI has no liability to anyone who relies on NCI measures and specifications or data reflective of performance under such measures and specifications.

Ad.8 Additional Information/Comments: We sincerely thank the teams and committees at NQF for all the support and guidance in this initial submission process.