



ADVANCING
STATES



**Review of
Psychometric Testing
of the National Core
Indicators – Aging and
Disabilities Adult
Consumer Survey**

Contents

Summary 3

Introduction to Validity and Reliability 3

 Figure 1. A visual depiction of validity and reliability 3

 Table 1. Description of the different types of validity and reliability 4

Validity and Reliability of NCI-AD ACS 5

Conclusions 7

References 8

Summary

This resource provides an overview of the psychometric testing conducted to date on the National Core Indicators – Aging and Disabilities™ (NCI-AD™) Adult Consumer Survey (ACS). The NCI-AD ACS is an initiative designed to support states in measuring the quality of long-term services and supports delivered to older adults and individuals with physical disabilities. A collaborative effort between Human Services Research Institute (HSRI), ADvancing States, and participating states, NCI-AD ACS has collected data on quality of life and outcomes of service recipients since 2015.

Modeled after the success of the National Core Indicators – Intellectual and Developmental Disabilities program, NCI-AD brings an important value proposition through indicators that measure priorities in the field in terms of social wellbeing, community participation, and person-centered services. Over the last decade, 38 states have participated in one or more survey cycles of NCI-AD ACS. Although there are many commonalities in how services are delivered across states, there are also many differences. The heterogeneity of state policies around service delivery underscores the importance of ensuring validity and reliability of NCI-AD ACS. **This report is intended to summarize what is known about the validity and reliability of NCI-AD ACS** for researchers, policy makers, state agency personnel and more; for more information about the methodology of NCI-AD ACS, including details about administration and proxy respondents, please see our [current reports](#).

Introduction to Validity and Reliability

Figure 1. A visual depiction of validity and reliability



Validity and reliability are the two main ways to understand the quality of a measurement tool, such as a survey. Together, validity and reliability ensure that the NCI questions about community inclusion, access, and satisfaction are accurate and

consistent measures of people’s experiences. Figure 1 shows a common way to illustrate the differences between validity and reliability.

Validity is another term for accuracy. As the image above shows, both target A and target B have poor accuracy – they miss the bullseye every time. Reliability is another important component to having a good survey tool; reliability is another term for consistency. As target A shows, you can have consistency but still be inaccurate. **The goal for surveys is to look like target C – the survey needs to measure what it intends to measure and do so consistently across different people in different states and at different times of their life.**

There are different types of validity and reliability that are examined in different ways. The table below provides a description of each type, and examples of ways to examine the different types of validity and reliability.

Table 1. Description of the different types of validity and reliability

Type of validity/ reliability	Description	How to examine instruments/measures
Face validity	A subjective assessment of whether the survey measure what it claims to measure at “face value.”	Subject matter expert review; focus groups; cognitive testing; Delphi panels
Content validity	A subjective assessment of whether all aspects of a concept are included in the measure.	Subject matter expert review; focus groups; cognitive testing; Delphi panels
Construct validity	An assessment of whether the tool accurately measures the constructs or domains it intends to measure.	In addition to the subjective methods, you can use more objective tests to look at whether constructs correlate with other measures of that construct; you can also use factor analysis to see if questions related to the same construct group together.
Criterion validity	An assessment of how well a survey relates to similar instruments that are valid and reliable.	Test for correlation to confirm that the survey predicts outcomes captured in other instruments.
Inter-rater reliability (IRR)	An assessment of whether different interviewers or surveyors obtain similar results when interviewing the same individuals.	Test for agreement between two or more interviewers. Importantly, IRR estimates account for the amount of agreement that is expected by chance alone.

Type of validity/ reliability	Description	How to examine instruments/measures
Internal consistency reliability	As assessment of whether questions that measure similar concepts produce consistent results.	Test for correlation between items that are related.
Test-retest reliability	An assessment of whether survey responses remain consistent over time when there are no changes in life circumstances.	Administer the survey at two different time points (ex: 2 weeks apart) and measure correlation between survey results.

Validity and Reliability of NCI-AD ACS

HSRI and ADvancing States conducted a number of psychometric validation activities during the survey development and [pilot year 2014-2015](#). These activities are briefly described below.

During the survey development phase, **face and content validity** were established through numerous iterations of expert review and input from stakeholders. A Steering Committee comprised of Aging and Disability staff from 19 states was created; this committee provided feedback on the target population, constructs that the survey ought to target, survey design considerations, and administration methods. Leading disability organizations, researchers, and quality measurement experts also provided input at this stage.

In support of **criterion validity**, the team reviewed the NCI-IDD In-Person Survey tool and more than 15 quality of life tools to look at how other surveys measured constructs the steering committee had identified as priority constructs. From these tools, the team created a list of 120 indicators across 17 domains. The steering committee participated in a modified Delphi panel to rate the indicators in terms of importance. This exercise produced a list of 61 indicators across 17 domains. The NCI-AD team then drafted questions to measure these indicators. The draft tool went through multiple rounds of revisions.

The next step to establish validity of the tool was cognitive testing with older adults and people with physical disabilities. The primary purpose of cognitive testing is to investigate if respondents understand the question as they are intended and if they can provide accurate answers. Cognitive testing examines the question-response process, which is generally viewed as consisting of four stages:

- Comprehension (respondent interprets the question)
- Retrieval (respondent searches memory for relevant information)

- Judgment (respondent evaluates his/her response)
- Response (respondent actually provides information in the requested format).

Cognitive testing is performed by conducting in-depth, semi-structured interviews with a small number of respondents similar to those targeted in the survey. The interviews are designed to elicit respondents' thought processes when answering the tested question, specifically, how they understood a question and how they arrived at their answer.

Cognitive testing of the NCI-AD ACS was conducted in two rounds. In round 1, cognitive testing involved two focus groups of individuals receiving publicly funded LTSS services in Massachusetts through aging and disability programs. 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 next draft of the survey, which was cognitively tested with several other LTSS participants in Maine; for the Maine cognitive testing, in-home survey administration mirrored planned NCI-AD survey environments. Data from cognitive interviews were qualitative, and analysis of those interviews helped the project team revise questions that were prone to errors at one of the four question response process stages. Because cognitive testing was conducted in two stages, it was possible to evaluate whether the revisions of the questions had the intended effect and resulted in improvement.

Inter-rater reliability was evaluated during the Maine cognitive testing study. Following revisions from cognitive testing, a pilot administration of the NCI-AD ACS was conducted in Georgia. The pilot in Georgia involved a total of 50 interviews; 24 of these being shadowed by a second interviewer who recorded the respondents' answers at the same time as the primary interviewer. The recorded responses were then compared and the level of agreement analyzed. It was found that there was a high level of agreement between interviewers (Cohen's Kappa > .80).

After the first full year of NCI-AD ACS survey data collection in 2015-2016, **construct validity** was studied using [exploratory and confirmatory factor analyses](#). External researchers screened items and assigned items to constructs; subsequently, an exploratory factor analysis (EFA) allowed for the examination of item properties and a confirmatory factor analysis (CFA) confirmed items aligned with intended constructs.

Of 88 items on the 2015-2016 NCI-AD survey, 36 items were initially selected for the subsequent factor analyses. Excluded items were typically conditional (e.g., only asked if you answered an earlier question as "yes" or response options allowed for checking all the apply). After the EFA, researchers excluded items that "cross-loaded" (e.g., loaded on two factors with a difference in factor loading that was less than 0.1) or had low factor loadings (<0.3) on any factor. Researchers also explored item missingness, item discrimination ability, and **item reliability tests**. Finally, they consulted subject

matter experts on the relevance and specific items and domain categories to ensure **content validity**.

Older Adults. EFA/CFA identified three factors made up of 16 items: security (5 items), care experience (4 items), and self-determination (7 items). Cronbach's alpha values are a measure of **internal consistency reliability**. Higher values (with 1.0 being perfect correlation) indicate higher reliability. For each factor, the values were 0.60, 0.70, and 0.52, respectively.

Younger Adults. EFA/CFA identified three factors made up of 15 items: security (5 items), care experience (4 items), and self-determination (6 items). Factor composition was similar for younger and older adults for the domains of security and care experience. The Cronbach's alpha values for each factor were 0.63, 0.67, and 0.66, respectively.

As part of every NCI-AD ACS survey, interviewers give feedback on any survey items that were problematic in an interview. They identify any questions that were difficult to ask, any questions that the responders had difficulty understanding, and any questions where the response options were not appropriate or sufficient. **Their feedback continues to be incorporated into revisions to ensure ongoing validity.**

Conclusions

The process used for developing the NCI-AD ACS followed rigorous methods to enhance the validity of the survey tool. Subsequent testing of the NCI-AD ACS has demonstrated that the survey has acceptable reliability. As NCI-AD ACS has grown to more states and more participants, HSRI requires completion of standardized trainings to surveyors to ensure reliable survey administration. We plan to continue testing inter-rater reliability of the NCI-AD ACS as well as conduct concurrent and predictive validity studies in the coming years.

References

National Core Indicators. (2016). National Core Indicators Aging and Disability Adult Consumer Survey: 2015-216 Mid-Year Results. https://nci-ad.org/upload/reports/NCI-AD_2015-2016_Six_State_Mid-Year_Report_FINAL2.pdf

Shippee, T.P., Duan, Y., Baker, Z.G., Parikh, R., Bucy, T. & Jutkowitz, E. (2025). *Measuring Consumer-Reported Quality of Life Among Recipients of Publicly Funded Home- and Community-Based Services: Implications for Health Equity*. *J Aging Health*, 37(7-8): 468-478. doi:10.1177/08982643241267378