

## CBE ID

2967

## Title

Home and Community-Based Services (HCBS) Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Measures

## Project

Advanced Illness and Post-Acute Care

## E&M Committee Rationale/Justification

**Committee Final Vote:** CBE #2967 - HCBS CAHPS contains 19 individual measures. Per the policy on Instrument-based Clinical Quality Measures the CBE does not endorse survey instruments. Rather, the CBE reviews and endorses measures derived from survey instruments in which survey assessments are aggregated to an accountable entity. Thus, each of the 19 measures derived from the HCBS CAPHS survey instrument is reviewed and endorsed separately.

The committee voted to endorse 17 of the 19 measures with conditions and did not reach consensus on two measures (#17 - *Unmet need in toileting due to lack of help* and #18 - *Unmet need with household tasks due to lack of help*), which resulted in endorsement being removed for those two measures.

**Condition:** For each of the 17 measures that received an endorsed with conditions designation, the committee would like to see the following condition addressed when these measures are submitted for maintenance review:

- The developer will explore methodological strategies (e.g., weighting, sampling) to ensure that responses are representative.

## Endorsement Decisions:

1. **Scale Measure 1 - Staff are reliable and helpful:** Endorsed with Conditions
2. **Scale Measure 2 - Staff listen and communicate well:** Endorsed with Conditions
3. **Scale Measure 3 - Case manager is helpful:** Endorsed with Conditions
4. **Scale Measure 4 - Choosing the services that matter to you:** Endorsed with Conditions
5. **Scale Measure 5 - Transportation to medical appointments:** Endorsed with Conditions
6. **Scale Measure 6 - Personal safety and respect:** Endorsed with Conditions
7. **Scale Measure 7 - Planning your time and activities:** Endorsed with Conditions
8. **Global Rating Measure 1 - Global rating of personal assistance and behavioral health staff:** Endorsed with Conditions
9. **Global Rating Measure 2 - Global rating of homemaker:** Endorsed with Conditions
10. **Global Rating Measure 3 - Global rating of case manager:** Endorsed with Conditions
11. **Recommendation Measure 1 - Would recommend personal assistance/behavioral health staff to family and friends:** Endorsed with Conditions
12. **Recommendation Measure 2 - Would recommend homemaker to family and friends:** Endorsed with Conditions

13. **Recommendation Measure 3 - Would recommend case manager to family and friends:** Endorsed with Conditions
14. **Unmet Needs Measure 1 - Unmet need in dressing/bathing due to lack of help:** Endorsed with Conditions
15. **Unmet Needs Measure 2 - Unmet need in meal preparation/eating due to lack of help:** Endorsed with Conditions
16. **Unmet Needs Measure 3 - Unmet need in medication administration due to lack of help:** Endorsed with Conditions
17. **Unmet Needs Measure 4 - Unmet need in toileting due to lack of help:** Endorsement Removed due to No Consensus
18. **Unmet Needs Measure 5 - Unmet need with household tasks due to lack of help:** Endorsement Removed due to No Consensus
19. **Physical Safety Measure - Hit or hurt by staff:** Endorsed with Conditions

## Is Under Review

No

## Next Maintenance Cycle

Spring 2029

## Previous Endorsement Cycle

Spring 2024

## Steward

Centers for Medicare & Medicaid Services

## 1.0 New or Maintenance

Maintenance

## 1.3 Electronic Clinical Quality Measure (eCQM)

No

## 1.6 Measure Description

The *Home and Community-Based Services (HCBS) Consumer Assessment of Healthcare Providers and Systems (CAHPS)* Survey consists of 19 measures that assess the experiences of Medicaid participants' age 18 and older receiving long-term services and supports (LTSS). The measures report a case-mix adjusted top-box score. Measure scores are calculated based on participant responses to a cross-disability survey about their experiences with the LTSS they receive in the community, delivered through a Medicaid-funded HCBS program. The unit of analysis is the Medicaid entity (that is, states, HCBS program, or managed care plans [MCPs], depending on the state).

## 1.7 Composite Measure

No

## 1.7 Measure Type

---

Patient-reported Outcome Performance Measure (PRO-PM)

## 1.8 Level of Analysis

Health Plan, Population or Geographic Area

### 1.8a Population or Geographic Area Level of Analysis

State

## 1.9 Care Setting

Other

### 1.9b Other Care Setting

Home and community-based services

## 1.10 Measure Rationale

The information that is collected as part of the HCBS CAHPS Survey informs HCBS managed care plans and states about their performance on services that are highly valued by HCBS participants. The type of data that are collected when implementing the survey is not readily available through other measures and can be used to target areas of improvement where scores are lagging. As the measure becomes implemented HCBS plans and states will have the ability to monitor performance over time and base care interventions, in part, on the trends they see in responses.

## 1.11 Measure Webpage

<https://www.medicaid.gov/medicaid/quality-of-care/quality-of-care-performance-m...>

## 1.13 Data Dictionary

Attached

## 1.14 Numerator

The HCBS CAHPS Survey measures are created using top-box scoring. This refers to the percentage of respondents that give the most positive response. Details regarding the definition of the most positive response are noted below. HCBS service experience is measured in the following areas: Scale Measures Staff are reliable and helpful—Average proportion of respondents that gave the most positive response on 6 survey items. Staff listen and communicate well—Average proportion of respondents that gave the most positive response on 11 survey items. Case manager is helpful—Average proportion of respondents that gave the most positive response on 3 survey items. Choosing the services that matter to you—Average proportion of respondents that gave the most positive response on 2 survey items. Transportation to medical appointments—Average proportion of respondents that gave the most positive response on 3 survey items. Personal safety and respect—Average proportion of respondents that gave the most positive response on 3 survey items. Planning your time and activities—Average proportion of respondents that gave the most positive response on 6 survey items. Global Rating Measures Global rating of personal assistance and behavioral health staff—Proportion of respondents that gave the most positive response of 9 or 10 on a 0-10 scale. Global rating of homemaker—Proportion of respondents that gave the most positive response of 9 or 10 on a 0-10 scale. Global rating of case manager—Proportion of

respondents that gave the most positive response of 9 or 10 on a 0–10 scale. Recommendation Measures Would recommend personal assistance/behavioral health staff to family and friends—Proportion of respondents that gave the most positive response of Definitely Yes on a 1–4 scale (Definitely No, Probably No, Probably Yes, or Definitely Yes). Would recommend homemaker to family and friends—Proportion of respondents that gave the most positive response of Definitely Yes on a 1–4 scale (Definitely No, Probably No, Probably Yes, or Definitely Yes). Would recommend case manager to family and friends—Proportion of respondents that gave the most positive response of Definitely Yes on a 1–4 scale (Definitely No, Probably No, Probably Yes, or Definitely Yes). Unmet Needs Measures Unmet need in dressing/bathing due to lack of help—Top-box score on a Yes or No scale. Unmet need in meal preparation/eating due to lack of help—Top-box score on a Yes or No scale. Unmet need in medication administration due to lack of help—Top-box score on a Yes or No scale. Unmet need in toileting due to lack of help—Top-box score on a Yes or No scale. Unmet need with household tasks due to lack of help—Top-box score on a Yes or No scale. Physical Safety Measure Hit or hurt by staff—Top-box score on a Yes or No scale.

### 1.14a Numerator Details

The Home and Community-Based Services (HCBS) Consumer Assessment of Healthcare Providers and Systems (CAHPS<sup>®</sup>) Survey is used to create 19 measures that include scale, global rating, recommendation, unmet needs, and physical safety measures. The measures are derived using top-box scoring, which refers to the percentage of respondents that give the most positive response. Details regarding the definition of top-box scores are noted below.

The HCBS CAHPS uses both positively and negatively worded questions to capture a comprehensive view of the HCBS participant experience. Higher scores on negatively worded questions indicate greater dissatisfaction or concerns with the aspects of care or services addressed in the questions.

For purposes of testing of psychometric properties in the Scientific Acceptability section, negatively worded items were reverse coded where appropriate.

#### **To calculate the program-level scores:**

Score each item using the top box method; calculate case mix adjusted scores for each item within a program; and calculate means for the scale measures.

#### *Scale Measures:*

For each survey item, the top-box numerator is the number of respondents who selected the most positive response category.

#### Staff are reliable and helpful—Survey items 13, 14, 15, 19, 37, and 38

13: In the last 3 months, how often did {personal assistance/behavioral health staff} come to work on time?

14: In the last 3 months, how often did {personal assistance/behavioral health staff} work as long as they were supposed to?

---

15: In the last 3 months, when staff could not come to work on a day that they were scheduled, someone let you know that {personal assistance/behavioral health staff} could not come that day?

19: In the last 3 months, how often did {personal assistance/behavioral health staff} make sure you had enough personal privacy when you dressed, took a shower, or bathed?

37: In the last 3 months, how often did {homemakers} come to work on time?

38: In the last 3 months, how often did {homemakers} work as long as they were supposed to?

Staff listen and communicate well—Survey items 28, 29, 30, 31, 32, 33, 41, 42, 43, 44, and 45

28: In the last 3 months, how often did {personal assistance/behavioral health staff} treat you with courtesy and respect?

29: In the last 3 months, how often were the explanations {personal assistance/behavioral health staff} gave you hard to understand because of an accent or the way {personal assistance/behavioral health staff} spoke English?

30: In the last 3 months, how often did {personal assistance/behavioral health staff} treat you the way you wanted them to?

31: In the last 3 months, how often did {personal assistance/behavioral health staff} explain things in a way that was easy to understand?

32: In the last 3 months, how often did {personal assistance/behavioral health staff} listen carefully to you?

33: In the last 3 months, did you feel {personal assistance/behavioral health staff} knew what kind of help you needed with everyday activities, like getting ready in the morning, getting groceries, or going places in your community?

41: In the last 3 months, how often did {homemakers} treat you with courtesy and respect?

42: In the last 3 months, how often were the explanations {homemakers} gave you hard to understand because of an accent or the way the {homemakers} spoke English?

43: In the last 3 months, how often did {homemakers} treat you the way you wanted them to?

44: In the last 3 months, how often did {homemakers} listen carefully to you?

45: In the last 3 months, did you feel {homemakers} knew what kind of help you needed?

Case manager is helpful—Survey items 49, 51, and 53

49: In the last 3 months, could you contact this {case manager} when you needed to?

51: In the last 3 months, did this {case manager} work with you when you asked for help with getting or fixing equipment?

---

53: In the last 3 months, did this {case manager} work with you when you asked for help with getting other changes to your services?

Choosing the services that matter to you—Survey items 56 and 57

56: In the last 3 months, did your [program-specific term for “service plan”] include . . .

57: In the last 3 months, did you feel {personal assistance/behavioral health staff} knew what’s on your [program-specific term for “service plan”], including the things that are important to you?

Transportation to medical appointments—Survey items 59, 61, and 62

59: Medical appointments include seeing a doctor, a dentist, a therapist, or someone else who takes care of your health. In the last 3 months, how often did you have a way to get to your medical appointments?

61: In the last 3 months, were you able to get in and out of this ride easily?

62: In the last 3 months, how often did this ride arrive on time to pick you up?

Personal safety and respect—Survey items 64, 65, and 68

64: In the last 3 months, was there a person you could talk to if someone hurt you or did something to you that you didn’t like?

65: In the last 3 months, did any {personal assistance/behavioral health staff, homemakers, or your case managers} take your money or your things without asking you first?

68: In the last 3 months, did any {staff} yell, swear, or curse at you?

Planning your time and activities—Survey items 75, 77, 78, 79, 80, and 81

75: In the last 3 months, when you wanted to, how often could you get together with these family members who live nearby?

77: In the last 3 months, when you wanted to, how often could you get together with these friends who live nearby?

78: In the last 3 months, when you wanted to, how often could you do things in the community that you like?

79: In the last 3 months, did you need more help than you get from {personal assistance/behavioral health staff} to do things in your community?

80: In the last 3 months, did you take part in deciding what you do with your time each day?

81: In the last 3 months, did you take part in deciding when you do things each day—for example, deciding when you get up, eat, or go to bed?

*Global Ratings Measures:*

---

The numerator for each global measure includes the number of respondents who answered 9 or 10 for the item (on a scale of 0 to 10).

Global rating of personal assistance and behavioral health staff—Survey item 35

35: Using any number from 0 to 10, where 0 is the worst help from {personal assistance/behavioral health staff} possible and 10 is the best help from {personal assistance/behavioral health staff} possible, what number would you use to rate the help you get from {personal assistance/behavioral health staff}?

Global rating of homemaker—Survey item 46

46: Using any number from 0 to 10, where 0 is the worst help from {homemakers} possible and 10 is the best help from {homemakers} possible, what number would you use to rate the help you get from {homemakers}?

Global rating of case manager—Survey item 54

54: Using any number from 0 to 10, where 0 is the worst help from {case manager} possible and 10 is the best help from {case manager} possible, what number would you use to rate the help you get from {case manager}?

*Recommendation Measures:*

The numerator for each recommendation measure includes the number of respondents who answered *Definitely Yes* for the item (on a scale of *Definitely No*, *Probably No*, *Probably Yes*, or *Definitely Yes*). Item numbers and item text are listed below.

Would recommend personal assistance/behavioral health staff to family and friends—Survey item 36

36: Would you recommend the {personal assistance/behavioral health staff} who help you to your family and friends if they needed help with everyday activities? Would you say you recommend the {personal assistance/behavioral health staff}?

Would recommend homemaker to family and friends—Survey item 47

47: Would you recommend the {homemakers} who help you to your family and friends if they needed {program-specific term for homemaker services}? Would you say you recommend the {homemakers}?

Would recommend case manager to family and friends—Survey item 55

55: Would you recommend the {case manager} who helps you to your family and friends if they needed {program-specific term for case-management services}? Would you say you recommend the {case manager}?

*Unmet Needs Measures:*

---

The numerator for each unmet needs measure includes the number of respondents who answered *No* for that item (these items are reverse coded so that higher scores reflect a better experience). Item numbers and item text are listed below.

Unmet need in dressing/bathing due to lack of help—Survey item 18

18: In the last 3 months, was this because there were no {personal assistance/behavioral health staff} to help you?

Unmet need in meal preparation/eating due to lack of help—Survey item 22

22: In the last 3 months, was this because there were no {personal assistance/behavioral health staff} to help you?

Unmet need in medication administration due to lack of help—Survey item 25

25: In the last 3 months, was this because there were no {personal assistance/behavioral health staff} to help you?

Unmet need in toileting due to lack of help—Survey item 27

27: In the last 3 months, did you get all the help you needed with toileting from {personal assistance/behavioral health staff} when you needed it? (not reverse coded).

Unmet need with household tasks due to lack of help—Survey item 40

40: In the last 3 months, was this because there were no {homemakers} to help you?

*Physical Safety Measure:*

The numerator for the following physical safety measure includes the number of respondents who answered *No* for this item (this item is reverse coded so that higher scores reflect a better experience). The item number and item text is listed below.

Hit or hurt by staff—Survey item 71

71: In the last 3 months, did any {staff} hit you or hurt you?

## **1.15 Denominator**

The denominator for all measures is the number of survey respondents. Individuals eligible for the HCBS CAHPS Survey include Medicaid participants who are age 18 and older in the sample period, and have received HCBS services for three months or longer. Eligibility is further determined using three cognitive screening items, administered during the interview: Does someone come into your home to help you? (Yes, No) How do they help you? What do you call them? Participants who are unable to answer these cognitive screening items are excluded. Some measures also have topic-specific screening items as well.

### **1.15a Denominator Details**

The denominator for all measures is the number of survey respondents. Individuals eligible for the HCBS CAHPS survey include Medicaid participants who are at least 18 years of age in the sample period and have received HCBS services for three months or longer.

While Medicaid programs provide a range of HCBS from different provider types (which vary by state) for participants with long-term services and supports needs, the proposed provider-related measures in this submission focus on the most common provider types for adults receiving Medicaid HCBS. These include personal assistance providers, behavioral health staff, homemakers, and case managers.

Personal care services and homemaker services typically involve assistance with activities of daily living, bathing, dressing, grooming, toileting, eating, mobility and instrumental activities of daily living, meal preparation, housework, laundry, food shopping. Case management is an integral component of Medicaid HCBS programs; the role of the case manager includes working with the participant to assess his or her need for services and supports and developing a person-centered care and service plan, referring individuals to needed services, monitoring service delivery, and responding to the individual's changing needs and circumstances.

Not all HCBS participants receive all services. Questions 4, 6, 8, 10, and 11 assess which services the participant receives. Participants are then eligible for different survey questions based on these responses.

These questions are:

4: In the last 3 months, did you get {program specific term for personal assistance} at home?

6: In the last 3 months, did you get {program specific term for behavioral health specialist services} at home?

8: In the last 3 months, did you get {program specific term for homemaker services} at home?

10: In the last 3 months, did the same people who help you with everyday activities also help you clean your home?

11: In the last 3 months, did you get help from {program specific term for case manager services} to help make sure that you had all the services you needed?

In addition to only including those eligible for the relevant survey questions based on a *Yes* response to one or more of the questions above, only individuals who provided a valid response to the individual survey items are included in each measure's denominator (i.e., participants for whom a *Don't Know* or *Refused*, or those for whom an unclear response was recorded, are not counted in a measure's denominator).

### **Use of Survey Data for Measure Testing (via Imputation of Personal Care Assistant and Homemaker Measure Data)**

In producing results for measures that rely on personal care assistant and homemaker data, the

measure developer built logic to denote places where the participant responded *Yes* to question 8 (“In the last 3 months, did you get {program specific term for homemaker services} at home?”) and question 10 (“In the last 3 months, did the same people who help you with everyday activities also help you clean your home?”). Through the imputation of data for personal care assistants and homemakers where values are missing, the measure developer could preserve responses and use as much data as feasible for calculating performance.

- i. All responses for questions 13 through 36 (i.e., the personal assistance questions) and questions 37 through 47 (i.e., the homemaker questions) were preserved irrespective of the instrument instructions (as displayed in **Exhibit 1**).
- ii. For instances where question 10 contained a value of *Yes* (or 01),
  1. The measure developer checked if there were valid responses for questions 13 and 47.
  2. If data were missing from these questions, the measure developer applied the following imputation strategy:
    - a. Impute values for questions\_13 through\_36 with responses from corresponding questions, 37 through\_47 (please see **Exhibit 2**); or
    - b. Impute values for questions 37 through 47 with responses from corresponding questions, 13 through 16 (please see **Exhibit 2**).

### *Scale Measures*

#### Scale Measure 1: Staff are reliable and helpful

13: The number of surveys completed by all those who responded *Yes* to screener 4, 6, or 10

14: The number of surveys completed by all those who responded *Yes* to screener 4,6, or 10

15: The number of surveys completed by all those who responded *Yes* to screener 4, 6, or 10

19: The number of surveys completed by all those who responded *Yes* to screener 4, 6, or 10

37: The number of surveys completed by all those who responded *Yes* to screener 8

38: The number of surveys completed by all those who responded *Yes* to screener 8

#### Scale Measure 2: Staff listen and communicate well

28: The number of surveys completed by all those who responded *Yes* to screener 4, 6, or 10

29: The number of surveys completed by all those who responded *Yes* to screener 4, 6, or 10

30: The number of surveys completed by all those who responded *Yes* to screener 4, 6, or 10

31: The number of surveys completed by all those who responded *Yes* to screener 4, 6, or 10

---

32: The number of surveys completed by all those who responded *Yes* to screener 4, 6, or 10

33: The number of surveys completed by all those who responded *Yes* to screener 4, 6, or 10

41: The number of surveys completed by all those who responded *Yes* to screener 8

42: The number of surveys completed by all those who responded *Yes* to screener 8

43: The number of surveys completed by all those who responded *Yes* to screener 8

44: The number of surveys completed by all those who responded *Yes* to screener 8

45: The number of surveys completed by all those who responded *Yes* to screener 8

Scale Measure 3: Case manager is helpful

49: The number of surveys completed by all those who responded *Yes* to 48

51: The number of surveys completed by all those who responded *Yes* to screener 11

53: The number of surveys completed by all those who responded *Yes* to screener 11

Scale Measure 4: Choosing the services that matter to you

56: The number of surveys completed by all those who responded *Yes* to screener 4, 6, 8, 10, or 11

57: The number of surveys completed by all those who responded *Yes* to screener 4, 6, 8, 10, or 11

Scale Measure 5: Transportation to medical appointments

59: The number of surveys completed by all those who responded *Yes* to screener 4, 6, 8, 10, or 11

61: The number of surveys completed by all those who responded *Yes* to screener 4, 6, 8, 10, or 11

62: The number of surveys completed by all those who responded *Yes* to screener 4, 6, 8, 10, or 11

Scale Measure 6: Personal safety and respect

64: The number of surveys completed by all those who responded *Yes* to screener 4, 6, 8, 10, or 11

65: The number of surveys completed by all those who responded *Yes* to screener 4, 6, 8, 10, or 11

68: The number of surveys completed by all those who responded *Yes* to screener 4, 6, 8, 10, or 11

Scale Measure 7: Planning your time and activities

75: The number of surveys completed by all those who responded *Yes* to screener 4, 6, 8, 10, or 11

77: The number of surveys completed by all those who responded *Yes* to screener 4, 6, 8, 10, or 11

78: The number of surveys completed by all those who responded *Yes* to screener 4, 6, 8, 10, or 11

79: The number of surveys completed by all those who responded *Yes* to screener 4, 6, or 10

80: The number of surveys completed by all those who responded *Yes* to screener 4, 6, 8, 10, or 11

81: The number of surveys completed by all those who responded *Yes* to screener 4, 6, 8, 10, or 11

*Global Rating Measures:*

Global rating of personal assistance and behavioral health staff

35: The number of surveys completed by all those who responded *Yes* to screener 4, 6, or 10

Global rating of homemaker

46: The number of surveys completed by all those who responded *Yes* to screener 8 and *No, Don't Know, Refused, or Unclear Response* to screener 10

Global rating of case manager

54: The number of surveys completed by all those who responded *Yes* to screener 11

*Recommendation Measures:*

Recommendation of personal assistance and behavioral health staff to family and friends

36: The number of surveys completed by all those who responded *Yes* to screener 4, 6, or 10

Recommendation of homemaker to family and friends

47: The number of surveys completed by all those who responded *Yes* to screener 8

Recommendation of case manager to family and friends

55: The number of surveys completed by all those who responded *Yes* to screener 11

*Unmet Needs Measures:*

Unmet need in dressing/bathing due to lack of help

18: The number of surveys completed by all those who responded *Yes* to 16 and *No* to 17

Unmet need in meal preparation/eating due to lack of help

22: The number of surveys completed by all those who responded *Yes* to 20 and *No* to 21

Unmet need in medication administration due to lack of help

25: The number of surveys completed by all those who responded *Yes* to 23 and *No* to 24

Unmet need in toileting due to lack of help

---

27: The number of surveys completed by all those who responded *Yes* to 26

Unmet need with household tasks due to lack of help

40: The number of surveys completed by all those who responded *No* to 39

*Physical Safety Measures:*

Hit or hurt by staff

71: The number of surveys completed by all those who responded *Yes* to screener 4, 6, 8, 10, or 11

## 1.15b Denominator Exclusions

No explicit exclusion criteria are specified, however the denominator is limited to participants who are at least 18 years of age in the sample period and have received HCBS services for three months or longer, as well as their proxies. During survey administration, additional exclusions include individuals for whom a qualifying response was not received for the cognitive screening questions mentioned in the denominator statement below.

## 1.15c Denominator Exclusions Details

Participants who do not provide an answer for one or more of the following cognitive screening items should be excluded:

1. Does someone come into your home to help you?
2. How do they help you?
3. What do you call them?

If a participant is unable to respond to these questions they may also use a proxy to respond for them. Individuals who are more likely to be good proxy respondents are: (a) those who are willing to respond on behalf of the participant; (b) unpaid caregivers, family members, friends, and neighbors; and (c) those who know the participant well enough that he or she is familiar with the services and supports the participant is receiving, having regular, ongoing contact with the participant. Examples of circumstances that increase the likelihood that someone has knowledge about the participant and their care situation include living with the participant, managing the participant's in-home care for a majority of the day, having regular conversations with the participant about the services they receive, in-person visits with the participant, and being present when services/supports are delivered. Individuals who are less likely to be good proxy respondents are: (a) those with paid responsibilities for providing services/supports to the participant, including family members and friends who are paid to help the participant; and (b) guardians or conservators whose only responsibility is to oversee the participant's finances. Due to the nature of data being collected through CAHPS, individuals who are paid to deliver HCBS services are discouraged from acting as a proxy.

## 1.16 Type of Score

Other

## 1.16a Other Scoring Method

Case-mix adjusted top-box score.

## 1.17 Measure Score Interpretation

Better performance = Higher score

## 1.18 Calculation of Measure Score

Details for calculation of the 19 measures based on HCBS CAHPS Survey responses are found in *Technical Assistance Guide for Analyzing Data from the Consumer Assessment of Healthcare Providers and Systems Home and Community-Based Services Survey*, which can be retrieved from <https://www.medicaid.gov/sites/default/files/2021-09/hcbshaps-appk-dat...>

In general, numerators are based on top box scores meaning that they reflect the number of eligible respondents who give the most positive response for each survey item.

The denominator for all measures reflects completed surveys from respondents eligible for the HCBS CAHPS Survey for the underlying survey item(s). For certain measures (e.g., scale measures that are a linear combination of multiple survey item responses), the denominator for a measure represents the denominator for each of the underlying survey item(s).

## 1.19 Measure Stratification Details

The intended primary unit of analysis is the Medicaid HCBS program. However, states may wish to stratify by sub-state agencies such as counties or regional entities with program operational and budgetary authority. In some instances, a state may wish to stratify by case-management agency as well, given they are typically viewed as having substantial responsibility for developing beneficiary service and support plans as well as monitoring whether the service or support plan addresses the person's needs and meet their goals.

States are increasingly moving users of Medicaid long-term services and supports (LTSS), including HCBS, into managed-care arrangements (typically referred to as Managed Long-Term Services and Supports [MLTSS]) where the managed care plan is the primary accountable entity for ensuring HCBS participant's health, welfare, and quality of life. As such, we also anticipate some states may want to stratify based on managed care plan results.

## 1.20 Types of Data Sources

Patient-Reported Data and/or Survey Data

### 1.21a Data Collection Tool URL(s)

<https://www.medicaid.gov/medicaid/quality-of-care/quality-of-care-performance-m...>

### 1.21b Attach Data Collection Tool(s)

---

[HCBS CAHPS Survey \(Version 1.0\).zip](#)

## 1.22 Proxy Responses

Yes

## 1.23 Survey Respondent

Patient, Family or Other Caregiver

## 1.24 Data Collection and Response Rate

*Proxy*, here, is defined as anyone who provided help to the participant completing the Survey. The measure developer allows proxy responses when completing the Survey.

The unit of analysis in the HCBS CAHPS Survey is typically the *accountable entity*, which refers to the entity responsible for managing and overseeing a specific HCBS program within a state. This entity could be a Medicaid agency, a non-Medicaid state agency (e.g., a department of aging), a county, or managed care plans via MLTSS programs. The choice of the unit of analysis, whether it is conducted at the program level or the managed care plan level, depends on the objectives of the data analysis and the characteristics of the available data (e.g., survey design, sample size, and variables).

### Response Rates

There is no minimum response rate for reporting any of the HCBS CAHPS measures. When the Survey was field tested, the response rate was 22.0 percent; this ranged from 9.8 percent to 31.1 percent for the different HCBS programs. Some states may expect a higher response rate because of better outreach, upfront communications, and use of proxies.

The measure developer calculated the response rate using the American Association for Public Opinion Research (AAPOR) response rate number 3:

$$I/((I+P) + (R+NC+O) + e(UH+UO))$$

Where:

I=Complete interviews (3,226)

P=Partial interviews (33)

R=Refusals and breakoffs (2,442)

NC=Non-Contact (3,014)

O=Other (3,200)

UH=Unknown household (3,868)

UO=Unknown other (123)

e=Estimated proportion of cases of unknown eligibility that are eligible (0.68)

AAPOR defines several options for calculating response rate. Based on the measure developer's sampling approach, the formula that is most appropriate for these data was response rate number 3 (described here: <http://aapor.org/wp-content/uploads/2023/05/Standards-Definitions-10th-...> ). The response rate is the number of completed surveys divided by the number of eligible sampled individuals. Households with nonworking or wrong numbers are excluded from the denominator. In some cases, eligibility cannot be determined. For these participants, response rate number 3 adjusts the response rate assuming that the rate of response for undetermined households would be the same as the response rate where eligibility could be determined. This is shown in the formula where the number of unknowns (UH+UO) is multiplied by the estimated proportion of cases of unknown eligibility that are eligible (e). The result is a slight upward adjustment of the response rate. Thus, the overall response rate was 21.1 percent (22.3 percent via in person and 20.9 percent via phone).

## 1.25 Data Source Details

Data are derived from participant responses to the HCBS CAHPS Survey, funded by the Centers for Medicare & Medicaid Services (CMS). The HCBS CAHPS instrument is available at <https://www.medicare.gov/medicaid/quality-of-care/quality-of-care-perfo...>;

## 1.26 Minimum Sample Size

To determine the size of the sample, each state considers the effective sample size and response rates (from the field test or the methodology described above). The effective sample size is the number of completed responses needed to obtain a reasonable level of reliability.

The measure developer conducted a pilot test and a field test of the measures with 26 Medicaid HCBS programs across 10 states from October 2013 to March 2015. Results suggest that the effective sample size should be 400 people per stratum (with smaller programs including the entire census). From field test data, the response rate was 22.0 percent; this ranged from 9.8 percent to 31.1 percent for HCBS programs and modes of administration. Some states expect a higher response rate in future administrations because of better outreach or pre-survey communications with potential respondents, as well as use of proxies. Estimated response rates can be adjusted to incorporate these additional considerations.

## 2.1 Attach Logic Model

[CBE 2967 Logic Model.pdf](#)

## 2.2 Evidence of Measure Importance

The measures in this submission focus on older adults, persons with physical disabilities, persons with intellectual or developmental disabilities, persons with acquired brain injury, and persons with mental health or substance use disorders who receive Medicaid-funded HCBS. The Medicaid population with disabilities is, by definition, a population with substantially limited economic

resources. Consistent with Medicaid status, adults with disability have a higher poverty rate than those without disability (aged 18 to 64: 24.8 percent versus 10.1 percent, respectively; aged 65 years and older: 15.1 percent versus 8.9 percent, respectively United States Census Bureau, 2022a)]. In addition, United States working-age adults (aged 18 to 64) with a disability have a lower employment rate than their non-disabled peers (46.6 percent versus 81.5 percent (United States Census Bureau, 2022b]).

For racial and ethnic disparities, persons who identify as Black or African-American, Hispanic, and American Indian or Alaskan Native have a higher prevalence of disabilities in self-care and independent living than does the total United States adult population with these types of disabilities; these types of disabilities mirror those that beneficiaries in Medicaid HCBS programs tend to exhibit. In the United States, 3.6 percent of the adult population has self-care disabilities and 7.2 percent have independent-living disabilities, respectively. This contrasts to Black or African-American individuals, who have a respective prevalence of 5.0 percent and 8.6 percent; Hispanics have a prevalence of 5.2 percent and 8.3 percent, respectively; and American Indians or Alaskan Natives have a prevalence of 7.1 percent and 13.7 percent, respectively (Centers for Disease Control and Prevention, 2021).

Safety is a major concern for programs serving people with disabilities, who experience higher rates of violent crime victimization. The rate of victimization from violent crime for the United States population without disabilities is 12.4 per 1,000 population. For persons with disabilities of the type served in HCBS programs (e.g., disabilities in self-care and independent living), the rates are 45.4 per 1,000 and 43.6 per 1,000 population, respectively. Of most relevance to the safety-related measure in this submission are statistics on victimization from abuse by paid caregivers; the Department of Justice's estimates do not identify paid caregivers as a category of perpetrator, however (Harrell, 2021).

#### *References:*

United States Census Bureau. (2022a). *American Community Survey, 1-year estimates, table B18130*. <https://data.census.gov>

United States Census Bureau. (2022b). *American Community Survey, 1-year estimates, table B18120*. <https://data.census.gov>

Center for Disease Control and Prevention. (2021). *Online disability and health data system Behavioral Risk Factor Surveillance System (BRFSS)*. <https://dhds.cdc.gov/>

Harrell, E. (2021). *Crime against persons with disabilities, 2009–2019—statistical tables*. <http://www.bjs.gov/content/pub/pdf/capd0913st.pdf>

## **2.4 Performance Gap**

The distribution of performance scores for the 19 HCBS CAHPS measures and distribution of respondents by measure are presented in **Exhibit 1** and **Exhibit 2** within the performance gap attachment. Scores for most measures demonstrate room for improvement, with median scores as low as 44.8 percent (unmet needs: *Sufficient staff to help you with meals*). Four measures

demonstrate high performance (i.e., the median performance score is above 90 percent), which include the scale measures—*Case manager is helpful* and *Personal safety and respect*, the unmet needs measure—*Sufficient staff to help you with toileting*, and the personal safety measure—*Do any staff hit or hurt you?* These performance scores, however, represent survey responses from the 20 participating programs included in the measure developer’s testing effort and may not be representative of the total eligible population who have not been surveyed. The median performance score is reported to reflect central tendency where outliers may exist. Performance by participant characteristics were examined to identify performance gaps in respective areas.

The median entity-level number of responses per measure (shown in **Exhibit 2**) ranges from 84.5 (for the *Recommendation of case manager* measure) to 123.0 (for the *Staff are reliable and helpful* measure), which includes the global ratings, recommendation, and single-item physical safety measure. The unit of analysis in the HCBS CAHPS Survey is typically the accountable entity, which refers to the entity responsible for managing and overseeing a specific HCBS program within a state. This entity could be a Medicaid agency, a non-Medicaid state agency (e.g., a department of aging), a county, or managed care plans via MLTSS programs. The choice of the unit of analysis, whether it is conducted at the program level or the managed care plan level, depends on the objectives of the data analysis and the characteristics of the available data (e.g., survey design, sample size, and variables). The five unmet needs measures had a much lower response rate, as the screening questions limit responses to those participants who indicated a need for the individual service being assessed. The median entity-level number of responses for the unmet needs measures ranged from 1 (for the *Sufficient staff to help you with meals* measure) to 56 (for the *Sufficient staff to help you with toileting* measure).

**Exhibit 3** through **Exhibit 9**, which appear in the performance gap attachment, examine performance scores for the 19 HCBS CAHPS measures for several participant characteristics (i.e., age band, gender, race, ethnicity, education level, language, and living arrangement). Chi-square statistics were calculated to determine whether differences in performance scores based on these characteristics were statistically significant.

Performance-score differences in education level (displayed in **Exhibit 7**) and race (displayed in **Exhibit 5**) were found to be significant at the 95-percent level for 14 of the 19 measures. Significant differences in performance were also found based on participant living arrangement (for 12 measures, as displayed in **Exhibit 9**), age band (for 10 measures, as displayed in **Exhibit 3**), language (for 8 measures, as displayed in **Exhibit 8**), and gender (for five measures, as displayed in **Exhibit 4**).

**Exhibit 10**, in the attachment, shows change in entity-level mean scores between survey years 2022 and 2023; these data show improvement for 12 of the 19 measures, ranging in magnitude from 1.1 percent (for the *Choosing the services that matter to you* measure) to 29.9 percent (for the *Sufficient staff to help dress, shower or bathe* measure). Data prior to the 2022 survey year were not included due to instability in collection procedures and responses provided during the COVID-19 public health emergency.

## 2.4a Attach Performance Gap Results

[CBE 2967 Performance Gap Tables.pdf](#)

---

## 2.6 Meaningfulness to Target Population

The HCBS CAHPS survey collects information from participants about the extent to which the services they receive are person-centered, prioritize what is important to them, and the extent to which participants can direct and control their own plan and how services are delivered. This is critically important feedback for states and programs, helping them see how well their efforts on adopting person-centered practices and approaches in service planning and delivery are working and what kind of a difference this is making in participants' lives.

Pennsylvania has used the HCBS CAHPS survey to inform quality measurement for five years. Pennsylvania reports that they use the survey to oversee the work that is being performed in real-time and allows them to get real-time feedback from participants. Pennsylvania also encourages survey implementors to review the survey results and data to explore various strengths and weaknesses. In 2021, Pennsylvania reports that they compared 2018, 2019, and 2020 survey results which helped them understand how well their program has been implemented, especially regarding participants' overall experience.

Connecticut has used the HCBS CAHPS survey to inform quality measurement for six years. Connecticut reports that the HCBS CAHPS survey provides the ability to compare participant experiences in the same domains across different waivers. Connecticut reports that comparison of program results helps highlight which providers are delivering the highest quality of service to participants. HCBS participants in Connecticut expressed appreciation for their waiver services and commented on the positive impact of the waiver programs on their lives (UConn Health Center of Aging, 2023).

### References:

Pennsylvania Department of Human Services. (2022, February 7). *2021 HCBS CAHPS statewide survey results* [PowerPoint slides]. PA providers. [https://www.paproviders.org/wp-content/uploads/2022/02/2021-HCBS\\_CAHPS-...](https://www.paproviders.org/wp-content/uploads/2022/02/2021-HCBS_CAHPS-...)

UConn Health Center of Aging. (2023). *Consumer Assessment of Healthcare Providers and Systems Home and Community-Based Services (HCBS CAHPS) survey results: Connecticut HCBS programs*. <https://health.uconn.edu/aging/wp-content/uploads/sites/102/2024/02/DSS...>

## 3.1 Contributions Towards Closing Care Gaps

As shown in **Exhibits 3** through **Exhibit 9**, as well as in **Exhibit 12** and **Exhibit 13**, some potential social risk factors were examined to identify performance gaps. These factors include age bands, gender, race, ethnicity, language spoken at home, education level, living arrangement, and health status. Differences in performance have been identified, which demonstrate an opportunity for improving health equity based on these risk factors.

## 4.1 Feasibility Assessment

Data Collection:

- Despite a substantial amount of training and an extensive guide provided to survey vendors, all did not follow the data collection instructions exactly. These aspects can be reinforced when reviewing and modifying the materials for future administrations.
- In addition, implementers will need to be thoroughly educated about skip patterns in the HCBS CAHPS survey instrument, applicability of questions to their programs, and how to explain this to data collectors and survey programmers (who will need to take these patterns into effect when analyzing the data). Some of these skip patterns may be adapted to specific states, in which case additional work will be required with survey vendors (e.g., to explain why the skip patterns were adapted and conduct additional review of the field disks to ensure the surveys were appropriately adapted).
- It will be important for states to provide clear specifications about the nature of the work and realistic information about the context in which vendors will need to work. This is especially critical if they decide to use a survey vendor that is not familiar with the data collection instrument or HCBS populations.

#### Sampling:

- We recommend screening the sample for deceased individuals to the greatest extent possible.

#### Response Rate:

- Many participants in Medicaid HCBS programs have guardians from whom consent for the person's participation in a survey must be secured. For many states, this information is not centrally or readily available, or not updated. Accessing this information prior to contact will help increase participation.
- The AAPOR response rates considers individuals who are deceased or who are physically or mentally unable to respond as eligible respondents resulting in lower response rates. An alternative is to calculate a response rate that does not include such individuals as eligible respondents.
- To avoid alarming potential survey participants and to enhance the recruitment process, any pre-notification letters to the participant should clearly identify the primary survey vendor.
- Programs should employ additional strategies for recruiting challenging populations, including using proxies. Additional outreach can involve case/care managers, or states might enlist advocacy groups to communicate to beneficiaries the importance of participating in the survey.

While field test response rates were less than optimal --22.0 percent, on average, they ranged from a high of 27.7 percent for participants from programs serving older adults, to a low of 9.8 percent for respondents in programs serving those with intellectual/developmental disabilities

(I/DD). The measure developer is confident that, as state use increases, the response rate will increase. Our confidence is based on: 1) the use of proxy respondents; 2) enhanced program-specific recruitment efforts; and 3) historical experience of response rates associated with a long-standing survey targeting people with I/DD. Taken together, these provide evidence of the feasibility of achieving sample sizes sufficient to discern variation and performance across programs. Each of these reasons for expecting increased response rates in the future is addressed in more detail below.

1. **Use Of Proxy Respondents:** A larger proportion of beneficiaries enrolled in programs serving people with I/DD tend to have guardians. It was our experience in the field test that guardians tended to act as gate keepers, refusing access to the participants. Many of these guardians said that the participant was not able to complete the survey but that as their guardian they wished to do so. In the beginning of the field test, proxy respondents were not allowed; at the time the CAHPS Consortium did not allow proxy respondents and since CMS was seeking a CAHPS trademark for the survey, the survey team did not allow proxy respondents. However, as field data started coming in, survey vendors began reporting that interviewers were allowing proxy respondents. Subsequently, the measure developer decided to allow and document proxy respondents so there would have opportunity to assess their contributions. However, because proxy respondents were allowed only beginning in September 2014 and data collection occurred over the period of July 2014 through February 2015, it is not possible to make definitive statements about the effect of proxy respondents on the response rate. That said, there is suggestive evidence that the response rate for the I/DD population may be increased if proxy respondents are allowed. Our analyses show that if proxy responses for persons with I/DD are counted, the number of respondents increases by approximately 100 percent.

The information below provides some additional information on proxy use during the field test for all populations, including information pertinent to proxy respondents for surveys targeting people with I/DD. Proxy respondents were most prevalent in programs serving people with I/DD. Nearly half of respondents for I/DD programs were proxy respondents whereas the proxy respondent rate was substantially less in programs serving people with other disabilities.

#### Proxy Respondents in the EoC Survey Field Test

Population	Proxy Complete (N)	Percent Proxy Complete Surveys	Range of State % Proxy Complete
I/DD	192	49.9%	36.1% - 85.7%
Older Adults	414	20.2%	5.2% - 36.6%
Acquired Brain Injury	53	20.7%	5.7% - 39.3%
Behavioral Health and SUD	8	2.6%	0.0% - 5.1%

---

Overall 667 22.2% 0.0% - 85.7%

If proxy respondents are allowed in future administrations of the survey one would expect an increase in response rates for the I/DD population. As noted above, we found that guardians, family members and caregivers acted as gatekeepers, not allowing access to the participant. For example, 14 percent of all eligible I/DD sample members had a guardian who did not allow access to the participant because they were either “physically or mentally incompetent” (an AAPOR category for non-response). Converting at least some gate-keeper guardians to proxy respondents in the future should increase response rates substantially.

2. Enhanced Program-Specific Recruitment Efforts: Since the conclusion of the field test data collection, the TEFT Demonstration state grantees have identified improvements that they intend to implement for the next round of data collection which they are conducting themselves. It is expected that some of these enhancements will result in improved response rates. They include:

- Ensuring that pre-notification letters originate from the state agency operating the HCBS program being surveyed so that those receiving the letter have familiarity with the letterhead.
- Ensuring that participant contact information is accurate by requiring that case managers verify participant and guardian contact information for persons sampled.
- Ensuring that survey vendors have experience and specialized qualifications with the populations being surveyed so they are sensitized to particular considerations in interacting with people with certain types of disability. This is likely to increase rapport and result in improved recruitment.
- Targeting survey mode to persons/groups more likely to respond to a certain mode (rather than randomization to mode as happened in the field test).
- Conducting outreach to relevant stakeholders about the survey. This includes case managers and providers so they can encourage beneficiaries to participate when they receive inquiries from sampled members about the legitimacy of the survey. It may also include family and caregiver support groups. Stakeholders are more likely to encourage survey participation if they understand who is sponsoring the survey, its purpose and benefits.
- Not fielding the survey during the winter holiday season.
- Not fielding the survey during winter months in colder climates due to the risk of inclement weather prohibiting travel.

3. Response Rates from An Other Survey of People with I/DD: Some of the TEFT Demonstration state grantees sponsor another survey - the National Core Indicators (NCI) survey

---

-- that elicits feedback from people with I/DD. Some state I/DD agencies have conducted the NCI repeatedly over many years. Consequently beneficiaries, family members and guardians are very familiar with the survey. Also, the NCI allows proxy respondents.

Four TEFT grantee states shared the response rates that they have attained in recent years for the adult NCI survey:

- Arizona: 87 percent response rate
- Kentucky: 94.5 percent response rate
- Colorado: 39 percent response rate
- Connecticut: In order to complete 400 surveys, they pull a sample of upwards of 1,000.

In addition to the information provided by the TEFT states, the National Association for Directors of Developmental Disability Services (NASDDDS), one of the sponsors of the NCI survey for people with I/DD, reports that for their face-to-face surveys: “Most states interview about 500 people to get the 400 sample size number (and most have to pull about 800 names to get the sample size).” ([http://www.nasddds.org/uploads/files/NCI\\_Description\\_and\\_Costs.pdf](http://www.nasddds.org/uploads/files/NCI_Description_and_Costs.pdf)). While the NCI project does not report average response rates, this information from the NASDDDS website indicates the feasibility of achieving much higher response rates for the ID/DD subgroup than realized in the Experience of Care survey field test.

Timing of Data Collection:

- States that experience snow/ice during the winter should be encouraged to schedule data collection in other seasons.

Administration time:

Prospectively, for the PRA package submitted to OMB, the research team estimated 30 minutes per administration. This estimate was based on the length of the survey and CMS’s experience with previous CAHPS surveys of comparable length that were fielded with a similar, although not identical, population – the long-stay nursing home resident CAHPS. The nursing home CAHPS reported an administration time of 20 minutes on average. Since there are more items in the CAHPS Home and Community-Based Services Survey than the Nursing Home CAHPS (96 items versus 45 items), the research team estimated 30 minutes for the PRA package.

Retrospective analysis post field test showed that on average, respondents answered 51 (out of 96 items). Skip patterns account for the majority of the discrepancy between total number of items and number answered. General guidance from a survey expert is that it takes a person can answer approximately 4 items per minute. This estimate would put the time to administer the HCBS CAHPS in the vicinity of 13 minutes. Although field test data are not available on length of time to administer, we can safely assume that survey administration, on average, did not exceed 30 minutes based on these other sources.

Routine use of satisfaction/experience surveys in HCBS programs:

State Medicaid programs have been administering surveys to their HCBS participants for many years. Many are home-grown and state-specific, others were developed with an intended wider audience; only some of these surveys have undergone limited testing. Participants and their caregivers are very accustomed to these surveys. We know that at least 46 states currently administer a satisfaction/experience survey to their beneficiaries in at least one HCBS program. However, these surveys are all disability-specific, unlike the HCBS CAHPS Survey, which was designed and tested as a cross-disability survey.

### **4.3 Feasibility Informed Final Measure**

Feasibility was assessed when the HCBS CAHPS measures were specified; these data were reviewed by the consensus-based entity in 2016. Information from that effort appears above.

### **4.4 Proprietary Information**

Not a proprietary measure and no proprietary components

#### **5.1.1 Data Used for Testing**

Participant responses to the HCBS CAHPS Survey for the measure year 2023 were used for all aspects of testing. These include responses from all agencies submitting data from the HCBS CAHPS survey in the 2023 measure year (which were collected in the 2022 calendar year) comprised of 5,799 completed surveys from 20 entities in seven states.

#### **5.1.2 Differences in Data**

None

#### **5.1.3 Characteristics of Measured Entities**

Measured entities include all agencies submitting data from the HCBS CAHPS survey in the 2023 measure year (which were collected in the 2022 calendar year). These included 5,799 completed surveys from 20 agencies in seven states. These 20 HCBS programs serve a wide array of participants, including those who are older adults, participants with a physical disability,

participants with an intellectual or developmental disability, participants with an acquired brain injury, and with a diagnosis of a mental health or substance use disorder condition.

See **Exhibit 11**, within the supplemental attachment, for more information on the program type(s) and number of completed Surveys, by state.

### 5.1.4 Characteristics of Units of the Eligible Population

The measure developer used 5,799 complete HCBS CAHPS Survey responses from 20 Medicaid HCBS entities in its analysis. The survey captured selected self-reported sociodemographic characteristics and health status (general and mental health) of the participants. **Exhibit 12** and **Exhibit 13** provide the count and distribution of responses.

### 5.2.1 Level(s) of Reliability Testing Conducted

Accountable entity level (i.e., measure score) (e.g., signal-to-noise analysis)

### 5.2.2 Method(s) of Reliability Testing

The measure developer tested measure reliability using multiple methods that address different aspects of a measure's reliability (e.g., consistency, repeatability), using all complete responses to test reliability of the measures. The unit of analysis (i.e., the level of data) used to calculate the statistical measures of reliability varied based on the measure and is indicated for each test.

*Internal Consistency Reliability (Unit of Analysis: Participant).* For scale (or multi-item) measures, the measure developer calculated Cronbach alpha (or Coefficient alpha) to assess internal consistency among individual items designed to measure a given construct. The Cronbach alpha measures the degree of interrelatedness between the individual items; it is calculated using participant-level data. A scale measure is considered reliable if Cronbach alpha is 0.70 or greater (Nunnally & Bernstein, 1994). **Exhibit 14**, available within the reliability attachment, shows Cronbach alpha values for the seven scalar measures.

*Inter-Unit Reliability (Unit of Analysis: Participant).* The measure developer estimated the inter-unit reliability for each measure to evaluate the extent to which the experiences of participants within an entity (i.e., an HCBS program or managed care plan) correlate with one another, when compared to the experience between entities. Consequently, it reflects the signal-to-noise ratio (i.e., the fraction of total variation that is due to signal, or the true variation in scores, across units). One of the primary purposes of the HCBS CAHPS measures is to detect differences among HCBS entities; thus, this ratio is a meaningful indicator of the extent to which the scale measures (and other survey items) accomplish this goal. Inter-unit reliability also indicates the level of reliability for a measure across participants. Inter-unit reliability represents a transformation of the F-statistic for testing differences among entities on a measure (where inter-unit reliability =  $\frac{F-1}{F}$ ). An inter-unit-reliability value close to 1 indicates that a majority of the variability is attributable to the entity; it is an indication that the ability of the item or scale measure to discriminate across programs is greater. Measures with reliability coefficients above 0.70 can be used for statistical analysis of entity-level comparisons<sup>1</sup> (Evensen et al., 2019; Keller et al., 2005; Nunnally, 1978). As the inter-unit-reliability value gets smaller, a larger sample is needed to

discriminate reliably across entities.

The measure developer also calculated intra-class correlation, which is defined as the between-unit variance minus the within-unit variance over the total variance, adjusted for the average number of participants per reporting unit (intra-class correlation (1,1)) (McGraw & Wong 1996). Based on acceptable best practices from a 2017 study, intra-class correlation estimates with values less than 0.50, between 0.51 and 0.75, between 0.76 and 0.90, and greater than 0.90 are indicative of poor, moderate, good, and excellent reliability, respectively (Koo & Li, 2017). Within **Exhibit 14**, the measure developer used 0.70 for consistency with other Consumer Assessment of Healthcare Providers and Systems analyses.

Inter-unit reliability provides reliability based on the sample size associated with the data available,<sup>2</sup> while intra-class correlation indicates the reliability of a measure for a single participant. The inter-unit reliability 200 values are the projected inter-unit reliability based on a sample of 200 responses per entity.<sup>3</sup> The measure developer estimated the inter-unit reliability 200 using the Spearman-Brown prophecy formula with the following formula:

where,

**Exhibit 14** shows the results of the inter-unit reliability, the inter-unit reliability 200, and the intra-class correlation analyses.

*Entity-Level Reliability (Unit of Analysis: Entity).* To test for reliability of items *within* an entity, the research team leveraged methods outlined in Adams's 2010 publication (Adams et al., 2010). The purpose of this signal-to-noise analysis is to measure the distribution of reliability scores across entities, rather than creating a single summary statistic (i.e., overall measure inter-unit reliability) to assess reliability.<sup>4</sup> For each measure,  $M$ , the measure developer estimated entity-level,  $i$ , inter-unit reliability (inter unit reliability <sub>$M_i$</sub> ), calculated as a signal-to-noise ratio, using the following formula:

For measures that are structured with a pass/fail (binary) response (e.g., proportion responding *Yes*) based on a single survey item (i.e., the global-rating, recommendation, unmet-need, and personal-safety metrics), the measure developer assumed that the pass rate (or measure score) is a binomial random variable conditional on the unit's true value that comes from the beta distribution ( $\beta$ ), where  $\alpha$  and  $\beta$  denote the shape parameters for the measure,  $M$ . The shape parameters for each measure was estimated using entity data that contain the number of pass ( $x$ ) values (e.g., the *Recommending personal assistant* measure), total participants for the measure ( $n$ ), and the *Betabin* SAS macro (Wakeling, n.d.). The variance of the beta-distributed variable, or the between-unit variability, for each measure,  $M$ , was then calculated based on estimated values of the shape parameter, which is described, as follows:

The within-unit variability for each measure,  $M$ , and unit,  $i$ , were estimated as

Within-unity variability  $\sigma^2$ , where

All composite measures (i.e., the scale measures [e.g., *Staff are reliable and helpful*]) are linear combinations of binary scale items.<sup>5</sup> The measure developer estimated the within-unit variability for each of these measures as variance of the linear combinations of binary variables.<sup>6</sup> The measure developer then estimated the between-unit variability, assuming a hierarchical normal-normal model, based on unit-level and estimated measure variance, using the following formula:<sup>7</sup>

#### References:

- Adams, J.L., Ateev, M., McGlynn, E.A. (2010). *Estimating reliability and misclassification in physician profiling*. RAND Corporation.  
[https://www.rand.org/pubs/technical\\_reports/TR863.html](https://www.rand.org/pubs/technical_reports/TR863.html)
- Evensen, C. T., Yost, K. J., Keller, S., Arora, N. K., Frentzel, E., Cowans, T., & Garfinkel, S. A. (2019). Development and testing of the CAHPS Cancer Care Survey. *Journal of oncology practice*, 15(11), e969–e978. <https://doi.org/10.1200/JOP.19.00039>
- Keller, S., O'Malley, A.J., Hays, R.D., Matthew R.A., Zaslavsky, A.M., Hepner, K.A., Cleary, P.D. (2005). Methods used to streamline the CAHPS Hospital Survey. *Health Serv Res*. 2005;40(6 Pt 2):2057-2077.<https://doi.org/10.1111/j.1475-6773.2005.00478.x>
- Koo, T.K., & Li, M.Y. (2017). A guideline of selecting and reporting intraclass correlation coefficients reliability research. *Journal of Chiropractic Medicine* 15(2):155-163.  
<https://doi.org/10.1016/j.jcm.2016.02.012>
- Nunnally, J.C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Nunnally, J.C. & Bernstein, I.H. (1994). The assessment of reliability. *Psychometric Theory*, 3, 248-292.
- McGraw, K.O., & Wong, S.P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30–46. <https://doi.org/10.1037/1082-989X.1.1.30>
- Portney, L.G., & Watkins, M.P. (2000). *Foundations of clinical research: applications to practice*. Prentice Hall.
- Wakeling, I. (n.d.). SAS macro for fitting beta-binomial models. *Qi statistics*. <http://www.qistats.co.uk/BetaBinomial.html>

#### Footnotes:

<sup>[1]</sup> Researchers (Nunnally, 1978; Keller et al., 2005; Evensen et. al, 2019) suggest that reliability values of 0.70 or greater are sufficient for most social research situations; CAHPS® data analyses (Instructions for Analyzing Data from CAHPS® Surveys, 2012) indicate the same.

<sup>2</sup> Nyce, Abery, and Ticha's observation that small sample sizes were not considered in earlier interpretations of inter unit reliability led to use of uniform projected sample size by the research team for this analysis.

<sup>3</sup>The measure developer conducted the signal to noise analyses in response to weaknesses of IUR, identified by Nyce, Abery, and Ticha, in *An Analysis of the HCBS CAHPS® Survey* (unpublished).

<sup>4</sup>As an example, the *Case manager is helpful* measure is estimated as total of proportion of respondents saying *Yes* and assigned a value of 1 to survey item (Q) 49, *Contact case manager when needed to, Yes* to Q51, *Case manager worked with you when asked for help getting or fixing equipment*, and *Yes* to Q53, *Case manager worked with you when you asked for help with getting other changes to your services*, divided by 3.

<sup>5</sup>Consider a scale measure, *SM*, as an example, which is a linear combination of two underlying items—*X* and *Y*—calculated as  $aX+bY$ , where *a* and *b* are constants. The variance of *SM* is estimated as  $\text{Var}(\text{SM})=a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X,Y)$ . The measure developer estimated variance and covariance using participant-level data and SAS procedures. For measures that include slightly different groupings of eligible responses (e.g., for the *Staff are reliable and helpful* measure), the measure developer estimated the covariance using data for common participants only.

<sup>6</sup>The measure developer applied the methods outlined by Adams (2010), using the SAS PROC MIXED program to estimate the between-unit variance, based on the unit-level data and variance estimate for the measure.

### 5.2.3 Reliability Testing Results

See reliability attachment.

#### 5.2.3a Attach Additional Reliability Testing Results

[CBE 2967 Reliability.pdf](#)

### 5.2.4 Interpretation of Reliability Results

**Exhibit 16**, within the reliability attachment, provides a summary of measure reliability, using testing results presented in **Exhibit 14** and **Exhibit 15** (also within the reliability attachment). The measure, •, indicates the measure met minimum reliability results (or has a value higher than the acceptable threshold). For entity-level reliability, the measure developer identified a measure as likely to be reliable if it had high reliability for more than 50 percent of the 20 entities for which the measure developer had data to perform the testing.

#### 5.3.1 Level(s) of Validity Testing Conducted

[Accountable entity level \(i.e., measure score\) \(e.g., criterion validity\)](#)

#### 5.3.3 Method(s) of Validity Testing

The HCBS CAHPS Survey measures are based on participant perception of care received within the HCBS space. There are no similar measures that exist to use in comparison testing to evaluate measure validity. Thus, the measure developer performed all tests of validity using available

CAHPS Survey response data from the 20 entities. To test validity, the measure developer performed four types of analyses:

- Face validity;
- Construct validity;
- Convergent validity; and
- Discriminant validity.

The measure developer used all complete responses to test validity of the HCBS CAHPS measures.

*Face Validity (Unit of Analysis: Expert Survey Respondent).* The measure developer assessed the single-item unmet needs measures and physical safety measure via a structured qualitative survey, using responses from a technical expert panel comprised of 10 individuals, including 3 participants, participant advocates, or caregivers.

*Construct Validity (Unit of Analysis: Participant).* Using participant-level data, the measure developer evaluated the extent to which the underlying items match the hypothesized factor structure (as demonstrated by their grouping for each scale measure).<sup>1</sup> Two tests were used for this assessment—confirmatory factor analyses and correlations between measure items.

Confirmatory Factor Analyses. Using participant-level data, the measure developer conducted confirmatory factor analyses, using the unweighted least squares method<sup>2</sup> to test for construct validity. **Exhibit 35** displays the model fit statistics, including absolute index standardized root mean square residual (which measures model fit without comparing with a baseline model; the cut-off for goodness-of-fit is the standardized root mean square residual is  $<0.08$ ), parsimony index-adjusted goodness-of-fit (which considers the model's complexity and incremental index; the cutoff for goodness-of-fit is adjusted goodness-of-fit is  $\geq 0.90$ ), the Bentler-Bonett normed fit index, or the Tucker-Lewis Index (which compares fit to a baseline model; the cutoff for goodness-of-fit is  $\geq 0.95$ ), and factor loadings for each item (Parry, n.d.). Factor loadings represent the correlation between each observed variable and its corresponding latent factor.

The magnitude of factor loadings indicates the strength of the relationship. Values closer to 1 suggest a stronger association, while values closer to 0 indicate a weaker association.

Correlation Analyses. The measure developer also examined correlations between the seven scale measures to determine if they measure different constructs. As the measures are based on participant response, it is expected that there will be some correlation between the items and, thus, the measures. For the measure constructs to be valid, however, all inter-scale measure correlations should be moderate to low (typically  $<0.80$ ) to indicate that these seven scale measures, although related, are not redundant.

*Criterion Validity (Unit of Analysis: Entity and Respondent).* The measure developer assessed the extent to which a measure construct is related to a relevant outcome of interest (i.e., the criterion), using the Point-Biserial Correlation (Kraemer, 2006) between scale measures and the relevant global rating and recommendation measures. For example, one can expect that the *Case manager is helpful* scale measure and the *Global rating of case manager* will be highly correlated, while the scale measures for *Staff are reliable and helpful* and *Staff listen and communicate well* will be highly correlated with the *Global rating of personal assistance staff* (and not the

instead of *Global rating of case manager*). Participant-level data were used to calculate these correlations; entity-level data were used to calculate entity-level measure correlations (i.e., Pearson correlation coefficient).

*Convergent Validity (Unit of Analysis: Measure)*. The measure developer evaluated the degree to which multiple measures of a single concept are interrelated. Strong correlations between the scores obtained from the new scale and the established measures provide evidence of convergent validity (Clark et al., 1995). The measure developer calculated Spearman correlation coefficients between scale-measure items using participant-level data. Note that item correlations calculated using participant-level data were estimated using pairwise deletion (i.e., the measure developer only used valid responses for corresponding items) to maximize the number of samples used to estimate the items' correlation.

*Discriminant Validity (Unit of Analysis: Participant)*. The measure developer assessed the degree to which items in a scale measure represent the measure (i.e., correlate) relative to other measures representing different concepts. The measure developer used a multi-trait analysis approach, proposed by Hays and Hayashi (1990), which extends the logic of multi-trait-multi-method analysis when only one method of evaluation is used (i.e., the HCBS CAHPS Survey instrument). The method includes estimating multi-trait-multi-item correlations between the item and a hypothesized "trait" (or measure); inter-correlations among scale measures; scale normality statistics; inter-correlations among items; and internal consistency reliability estimates for the scale measures. Based on recommendations in the Hays and Hayashi paper, item discrimination is supported if:

- The highest correlation in a row of the multi-trait-multi-item matrix is the correlation between the item and the "trait" that it is hypothesized to evaluate;
- The correlation is  $\geq 0.40$ ; and
- The correlation is significantly larger than the other correlations in the row.

The measure developer used the MULTI SAS program (Hays & Wang, 1992) for the multi-trait analysis.<sup>3</sup>

#### References:

- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7(3), 309-319. <https://doi.org/10.1037/1040-3590.7.3.309>
- Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling*, 16(4), 625-641. <https://doi.org/10.1080/10705510903203573>
- Hays, R.D., & Hayashi T. (1990). Beyond internal consistency reliability: Rationale and user's guide for multitrait scaling analysis program on the microcomputer. *Behavior Research Methods, Instruments, and Computers*, 22(2):167-175.
- Hays, R.D., & Wang, E. (1992). Multitrait scaling program: MULTI. *Proceedings of the Seventeenth Annual SAS Users Group International Conference*. [https://labs.dgsom.ucla.edu/hays/pages/programs\\_utilities](https://labs.dgsom.ucla.edu/hays/pages/programs_utilities)

Kraemer, H.C. (2006). Biserial correlation. In *Encyclopedia of Statistical Sciences* (eds S. Kotz, C.B. Read, N. Balakrishnan, B. Vidakovic and N.L. Johnson). <https://doi.org/10.1002/0471667196.ess0153.pub2>

Li, C. H. (2016). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods*; 21(3), 369–387. <https://doi.org/10.1037/met0000093>

Parry, S. (n.d.). Fit Indices commonly reported for CFA and SEM. Cornell Statistical Consulting Unit. [https://www.cscu.cornell.edu/news/Handouts/SEM\\_fit.pdf](https://www.cscu.cornell.edu/news/Handouts/SEM_fit.pdf)

#### Footnotes:

<sup>1</sup> Participant-level survey data can have item non-responses for multiple reasons, including participant understanding of question and willingness to respond. As factor analysis performs list-wise deletion for missing items (the purpose of the analysis is to test for validity of the measure construct), to maximize the sample size used for testing and mitigate impact of randomness inherent to survey data, the measure developer imputed missing responses.

The measure developer performed sensitivity analyses testing using two different types of imputation methods. For the first imputation method, data were subset for each scale measure if there were at least one non-missing item for the measure. The missing value for the participant (for the relevant scale measure item[s]) was imputed based on non-missing values for that item; doing so assumed that there was no pattern to missing values (i.e., that data were missing arbitrarily). Within SAS, *PROC MI* with *MCMC CHAIN=MULTIPLE* and *IMPUTE=FULL* options were used to estimate the missing item response. Detailed information on this SAS procedure and its syntax is available at <https://support.sas.com/documentation/onlinedoc/stat/141/mi.pdf>. For the second imputation method, the measure developer coded all non-response values to 99. This is an industry standard coding nomenclature for missing values for binary variables (i.e., 0/1 responses).

<sup>2</sup> Since items for these measures are ordinal (i.e., data that are ranked), the measure developer explored two methods recommended by Li (2016)—unweighted least squares and diagonally weighted least squares. Although the incremental and parsimony fits were slightly better using diagonally weighted least squares (when compared to unweighted least squares), evidence from the literature suggests using the unweighted least squares method, as it provides a more precise estimate (Ferreiro, 2009).

<sup>3</sup> Participant-level data can have item non-responses for many reasons. As the multi-trait analysis performs list-wise deletion for missing items, the measure developer imputed missing values to maximize the sample used for each measure. Data were subset for each scale measure if there were at least one non-missing item for the measure. The missing value for the participant (for the relevant scale measure item[s]) was imputed based on non-missing values for that item; doing so assumed that there was no pattern to missing values (i.e., that data were missing arbitrarily). Within SAS, *PROC MI* with *MCMC CHAIN=MULTIPLE* and *IMPUTE=FULL* options were used to estimate the missing item response. Detailed information on this SAS procedure and its syntax is available at <https://support.sas.com/documentation/onlinedoc/stat/141/mi.pdf>.

### 5.3.4 Validity Testing Results

*Face Validity.* Due to lack of empirical testing, a systematic assessment of face validity was included to provide evidence of face validity. Data for the five unmet needs measures and one personal safety measure were collected via a survey of 10 members of the measure developer's technical expert panel. Results appear in **Exhibit 17** through **Exhibit 34** (see validity attachment)

*Construct Validity.* **Exhibit 35** and **Exhibit 36** display estimates related to the two construct validity tests—confirmatory factor analysis and scale measure correlation. **Exhibit 35** presents the model fit statistics and factor loading from the confirmatory factor analysis. A majority of questions for each factor had loadings of 0.40 or greater; only one value fell below 0.20. Moreover, the questions with scores below 0.40 did not load onto multiple factors. **Exhibit 36** shows the correlation between the scale measures using participant-level data (all exhibits within validity attachment).

*Convergent Validity.* **Exhibit 38** through **Exhibit 44** (all within the validity attachment) display the inter-item correlation between items of same scale measure, used to assess for measure convergent validity. A moderate to high correlation between the items indicate that they capture a similar concept.

*Discriminant Validity.* **Exhibit 45**, contained within the validity attachment, displays the results of the multi-trait analysis for the scale measures to assess for measure discriminant validity.

#### 5.3.4a Attach Additional Validity Testing Results

[CBE 2967 Validity.pdf](#)

### 5.3.5 Interpretation of Validity Results

The measure developer performed validity testing of all measures using methods appropriate for each measure. **Exhibit 46**, within the validity attachment, presents a summary of the validity testing. A detailed description of results for each type of validity testing appears below the exhibit.

#### **Exhibit 46. Summary of Validity Testing**

*Face Validity for Single-Item Measures.* The measure developer conducted a systematic assessment of the single-item unmet needs measures and physical safety measure using a qualitative survey of members of its technical expert panel. Ten individuals responded to the survey. Responses provided feedback using a Likert scale with the five possible responses—*Strongly Agree*, *Agree*, *Undecided*, *Disagree*, and *Strongly Disagree*. Face validity was deemed to be high if more than 70 percent of responses were *Strongly Agree* or *Agree*; moderate if *Strongly Agree* or *Agree* fell between 50 and 70 percent; and low if the *Strongly Agree* and *Agree* responses were below 50 percent.

*Construct Validity for Scale Measures.* The measure developer used confirmatory factor analysis and correlation to test the hypothesized factor structure for the scale measures.

- Confirmatory Factor Analysis. The estimated model had a reasonably good fit, with a Standardized Root Mean Square of 0.05; Adjusted Goodness of Fit of 0.96, and Normed Fit

Index of 0.94;<sup>1</sup> these values indicate that the measure constructs are valid.

Ideally, the measure developer would seek a factor loading value between 0.40 and 1.00 to indicate good correlation between observed items and latent factors. For the scale measures, *Staff Listen and Communicate Well*, *Choosing the Services that Matter to You*, *Transportation to Medical Appointments*, and *Planning your Time and Activities*, a majority of item factor loadings fell within this range. For scale measures, *Staff are Reliable and Helpful*, *Case Manager is Helpful*, and *Personal Safety and Respect*, most item factor loadings fell below 0.4 or were greater than 1.0; these values may indicate multicollinearity. Scale items not falling within statistical thresholds are considered acceptable, however, based on face validity and general acceptance as being theoretically justified.

- **Correlation.** The inter-scale measure correlations were low (see **Exhibit 36**), with the highest correlation seen between the *Staff are reliable and helpful* and *Staff listen and communicate well* measures, at 0.45. These results indicate that the measure constructs do not overlap.

**Criterion Validity.** **Exhibit 36**, containing concurrent criterion validity analyses, show the following:

- The *Global rating of personal care assistant* and *Recommendation measure of personal care assistant* measures are moderately correlated with the *Staff are reliable and helpful* and *Staff listen and communicate well* scale measures (i.e., values were above 0.30). Additionally, these measures have relatively low correlation with other scale measures that define different measure constructs.
- The *Global rating of homemaker* and *Recommendation measure of homemaker* are moderately correlated with the *Staff are reliable and helpful* and *Staff listen and communicate well* scale measures (i.e., values were above 0.30). Additionally, these measures have relatively low correlation with other scale measures that define different measure constructs.
- The *Global rating of case manager* and *Recommendation measure of case manager* have low correlations with the *Case manager is helpful* (0.17 and 0.20, respectively) and *Choosing the services that matter to you* (0.26 and 0.20, respectively).
- The single-item physical safety measure had low correlation with the *Personal safety and respect* scale measure (i.e., value was 0.08).
- Low correlations are observed between the *Personal safety and respect* and *Case manager is helpful* scale measures when compared to the global rating and recommendation measures. There was significantly high correlation across all other items used in the scale measures, ratings, and recommendation measures for personal care assistant and behavioral health staff, homemakers, and case managers. These high correlations indicate strong validity of the data elements and the instrument, overall, in measuring HCBS participants' experience with their services.

**Convergent Validity.** At the participant level, nearly all of the intra-scale measure questions correlated significantly with one-other (see **Exhibit 38** through **Exhibit 45**). These data support the validity of the within-scale measure data elements in measuring the individual scale measures for HCBS participants.

**Discriminant Validity for Scale Measures.** For all items, in the multi-trait-multi-item matrix, the

correlation between the item and the “trait” that it is hypothesized to measure was the highest (i.e., the scaling success was 100 percent) and the correlation was 0.40 or above; collectively these data demonstrate item discrimination. *Staff are reliable and helpful* and *Staff listen and communicate well* met the criteria for reliability (as observed in the **Reliability Testing** section). While all other measures met the scaling criteria, they did not meet the reliability criteria. As observed for reliability testing, *Personal safety and respect* did not meet the Cronbach reliability criterion, as most participants reported the highest score for all the items (which resulted in very little variance).

The measure developer notes that, for correlations calculated using participant-level data, the calculations were based off of pairwise deletion of records where responses were missing. Thus, the item correlations are only based on data with responses for a subset of the responses used for the case manager question and other scale measure items. In contrast, the correlation between entity-level measures were calculated using the estimated measures for each entity and include all responses appropriate for inclusion in measure calculation for each entity.

#### Footnotes:

<sup>1</sup> The threshold for good model fit are: (i) SRMR of <0.08, (ii) AGFI of  $\geq 0.90$ , and (iii) NFI  $\geq 0.95$ .

<sup>2</sup> For example, nearly 92 percent of participants responded to questions captured in the *Staff are reliable and helpful* measure, compared to a 72-percent response rate for the *Case manager is helpful* measure (see **Exhibit 2**). Based on the HCBS CAHPS Survey skip logic and participant preference, not every participant answered every question. Approximately 4,200 participants responded to the *Staff come to work on time* (Q13) item and 1,300 participants responded to the *Case manager helped when asked for help with getting or fixing equipment* (Q51) question; of these, only 800 answered both questions.

### 5.3.2 Type of Accountable Entity Level Validity Testing Conducted (derived)

Empirical validity testing at the accountable entity-level (e.g., criterion validity, construct validity, known groups analysis), Systematic assessment of face validity of the measure’s performance score as an indicator of quality or resource use

#### 5.4.1 Methods Used to Address Risk Factors

Statistical risk adjustment model with risk factors

#### 5.4.2 Conceptual Model Rationale

The Medicaid HCBS population is diverse and includes many participants with social and/or functional status-related risk factors. Nearly two-thirds of participants are eligible for HCBS due to a disability, and around half are not white (MACPAC, 2018) (Peebles et al., 2017).<sup>1,2</sup> Participants are also generally low income and have lower levels of high school or college education. These factors place them at greater levels of risk for conscious and unconscious bias in the healthcare system, according to the National Institute on Minority Health and Health Disparities Research Framework (National Institute on Minority Health and Health Disparities, 2017).

The combination of individual factors listed above, in addition to the caregiver and community

aspects of HCBS, as well as the societal policies and structures that often marginalize HCBS participants, can all influence the quality of care that participants receive. The HCBS CAHPS Survey can monitor and evaluate these disparities, as the survey's data allow for stratified analyses on social risk factors (e.g., disability, race, ethnicity, gender, primary language, and education).

These factors are captured in the risk model below (see **Exhibit 47**, within the attachment).

#### References:

MACPAC. (2018). *Medicaid home- and community-based services: Characteristics and spending of high-cost users*. Medicaid and CHIP Payment and Access Commission. <https://www.macpac.gov/wp-content/uploads/2018/06/Medicaid-HCBS-Charact...>

National Institute on Minority Health and Health Disparities. (2017). *NIMHD research framework*. U.S. Department of Health & Human Services. <https://www.nimhd.nih.gov/research> Framework

Peebles, V., Kim, M., Bohl, A., Morales, N., Lipson, D. (2017). *HCBS claims analysis chartbook: Final report*. Medicaid and CHIP Payment and Access Commission. <https://www.macpac.gov/wp-content/uploads/2018/06/HCBS-Claims-Analysis-...>

### 5.4.2a Attach Conceptual Model

[CBE 2967 Risk Model.pdf](#)

### 5.4.3 Variable Distribution Across Measured Entities

See **Exhibit 12**, within the performance gap attachment, which details the number of responses by reported participant sociodemographic characteristics.

### 5.4.4 Risk/Case-Mix Adjustment Modeling and/or Stratification Results

The Centers for Medicare & Medicaid Services and the Agency for Healthcare Research and Quality encourage implementers of the HCBS CAHPS Survey to utilize the CAHPS Analysis Program, which is available at <https://www.ahrq.gov/cahps/surveys-guidance/helpful-resources/analysis/...> This program consists of seven SAS files, including test data sets, test programs, and a macro, which should remain unaltered. Use of the program facilitates consistent analysis of survey data, enabling implementers to adjust data, statistically, for valid comparisons across states and managed care programs. Special functions within the program cater to specific data features (e.g., unequal item weighting, sampling weight, stratifying estimates). Missing responses in survey data are addressed through imputation, with the program offering options for imputing missing values for adjuster variables. Users are advised to carefully consider the appropriateness of imputation for variables with missing values, ensuring imputation only where necessary.

Those who implement the HCBS CAHPS Survey and who wish to generate case-mix adjusted scores (e.g., mean or top-box scores) without utilizing the CAHPS Analysis Program can do so through multivariable analysis using their preferred software. The multivariate analysis generally

involves employing general linear models, including ordinary least squares regression. Adjusted scores are derived as predicted scores from such models. To perform this analysis, implementers of HCBS CAHPS should follow these steps:

- Begin by creating unadjusted scores for all single-item and scale measures to serve as the dependent variable in the analysis.
- Utilize the unit of analysis (e.g., HCBS program, managed care plan, case management agency) as the primary independent variable, with case-mix adjusters acting as covariates or control variables, and the dependent variable as the score of interest (e.g., global rating, scale score, item score).
- Calculate the least square means (marginal means) for each unit to determine whether each program's score deviates from the overall mean.
- Conduct a t-test to assess the statistical significance of differences between individual and overall means, excluding the entity of interest.

#### 5.4.5 Calibration and Discrimination

We calculated R-squared values across the three global rating and seven scale measures. The median adjusted R-squared was 0.05. The minimum was 0.02 and the maximum was 0.14. While these values may appear low, they demonstrate that most of the variance in scores between HCBS programs is not due to patient characteristics like age, sex, education, race, ethnicity, language, and self-rated mental and physical health. This does not mean that these variables play no role in explaining variation in performance, rather even after controlling for these variables there are still substantial differences in performance across HCBS programs.

#### 5.4.6 Interpretation of Risk/Case-mix Factor Findings

The goals of case-mix adjustment are to correct or remove the effects of individual participant characteristics, which may affect ratings at an entity level, and remove effects that might be considered spurious (i.e., that reflect something other than quality of care). To ensure consistency with the original package evaluated by the consensus-based entity in 2016, the measure developer maintained a similar approach to the methodology implemented for field testing eight years ago.

The variables for case-mix adjustment were determined based on the following three conditions:

- Case-mix variables reflect characteristics that are brought to the HCBS program by the participant (e.g., age, education, race); they are not traits that result from the participant's experience with, or assessment of, the HCBS program (e.g., number of visits with a case manager).
- Case-mix variables have reasonable correlation with measures or items within entities. Specifically, the approach to adjustment evaluates whether the variables have sufficient predictive power in relation to the outcomes (e.g., older adult participants give higher ratings of their care when compared to younger participants).
- There is variation between entities for these predictor variables, which is referred to as *heterogeneity*. One HCBS program, for example, may have participants that tend to be much younger than the population served by another HCBS program.

Individual characteristics (i.e., age, race, ethnicity, education, language, living arrangements, general health status, and mental health status) are defined by the Centers for Medicare & Medicaid Services and the Agency for Healthcare Research and Quality as having strong and consistent associations with consumer feedback in other Consumer Assessment of Healthcare Providers Surveys. Field testing for the 2016 HCBS CAHPS consensus-based-entity submission also identified several design characteristics—survey administration mode (i.e., in person, via phone), response option (i.e., standard, alternate), proxy status (i.e., whether another party completed the survey on behalf of the respondent), and assistance with the survey (i.e., whether another party helped the respondent complete the survey) as important factors. The survey design and administration data to which the measure developer currently has access only includes information on proxy and assistance status. Thus, the measure developer used these respondent and survey-design characteristics as potential case-mix adjusters for the HCBS CAHPS measures.

To complete case-mix selection and reporting, the measure developer followed four steps:

1. Select potential case-mix adjusters for each measure;
2. Estimate measure and case-mix adjuster heterogeneity;
3. Estimate predictive power of the selected adjusters; and
4. Estimate the impact of each adjuster.

To select potential case-mix adjusters for each measure, the measure developer used stepwise regression (i.e., a forward-selection method) to select the potential case-mix adjusters for each HCBS CAHPS measure. The stepwise regression analyses evaluated the strength of the relationship of each potential adjuster to the three global rating and six scale measures in separate models, in which each measure was regressed on all of the potential adjusters. In the stepwise regression models, the potential adjuster variables were added individually to the model. For a variable to remain in the model, its F-statistic had to be significant at the  $p < 0.10$  level. Adjuster variables selected in any of the models formed a core set of potential case-mix adjusters eligible for final selection.

To estimate measure and case-mix-adjuster heterogeneity, the measure developer evaluated the heterogeneity of outcome variables across entities—the ratio of between-entity to within-entity variance of the residuals when each variable was regressed on the entity in a random effects model. Heterogeneity of the predictor variables across entities was measured as the ratio of between-entity to within-entity variance of the residuals when each variable was regressed on all other potential case-mix adjusters in a random effects model, where the entity was included in the model as a random effect.

To estimate predictive power of the selected adjusters, the measure developer evaluated predictive power as the incremental amount of variance explained by the predictor (represented as the partial  $r^2 \times 1,000$ ) in stepwise regression analyses, controlling for the other potential case-mix adjusters.

To measure explanatory power, which considers both the predictive power of each potential adjuster and the heterogeneity of the adjusters across programs, the measure developer multiplied the predictive power by the adjuster heterogeneity factor.

Finally, the measure developer calculated the impact factor, which standardizes explanatory

power with respect to the overall variance in the outcome being assessed as explanatory power/outcome heterogeneity. Variables that had an impact factor >1.0 are considered as candidates for potential case-mix adjusters. The heterogeneity of the measures across entities, heterogeneity of the selected case-mix adjusters, predictive power of selected case-mix adjusters for each relevant measure, and whether the adjuster has potential impact are shown in **Exhibit 48** through **Exhibit 50** (within the risk adjustment attachment).

Based on the results displayed in **Exhibit 48** through **Exhibit 50**, case-mix adjusters appear to have a significant impact, depending on the measure type. Distribution of participant age varied significantly across entities, which demonstrates higher heterogeneity when compared to other case-mix adjusters; these variables were significant factors for most measures. Participant age bands, education, and race were important case-mix adjusters for measures evaluating homemaker and case-manager services. Though educational achievement status was not captured in the 2016 version of the HCBS CAHPS Survey and (and, thus, its testing for inclusion in case-mix adjustment), these variables have an impact on scale measures (e.g., *Staff are reliable and helpful*, *Staff listen and communicate*, and *Access to medical transportation*).

Race appears to have an impact on evaluation of the personal care assistant or behavioral health staff, homemaker, and case manager services, as well as several scale measures (e.g., *Staff listen and communicate*, *Case management*, and *Access to medical transportation*). Both sets of health status variables (for general health and mental health) have a significant impact on certain measures (e.g., *Staff reliable and helpful*, *Choosing services that matter to you*, *Personal safety*, and *Community inclusion*). Use of a proxy by participants has a significant impact on multiple measures (e.g., *Staff reliable and helpful*, *Choosing services that matter to you*, and *Personal safety*).

#### 5.4.7 Final Approach to Address Risk Factors

Statistical risk adjustment model with risk factors

#### 6.1.1 Current Status

In use

#### 6.1.3 Current Use(s)

Quality Improvement (Internal to the specific organization)

#### 6.1.3 Program Details

Name of the program and sponsor

Home and Community-Based Services Measures

URL of the program

<https://www.medicaid.gov/medicaid/home-community-based-services/index.html>

Purpose of the program

See applicable level of analysis and care setting (below).

Geographic area and percentage of accountable entities and patients included

See applicable level of analysis and care setting (below).

Applicable level of analysis and care setting

*Purpose:* The Centers for Medicare & Medicaid Services provides opportunities for Medicaid

beneficiaries to receive services in their own home or community rather than institutions or other isolated settings. To evaluate these services, measures like HCBS CAHPS are made available to states and managed care plans to evaluate the experience of care for HCBS participants.

*Geographic area and percentage of accountable entities and patients included:* The quality improvement measures are calculated at the state or program level. Twelve states—Arizona, Connecticut, Florida, Hawaii, Kansas, Louisiana, Maine, Michigan, Missouri, North Carolina, Pennsylvania, and West Virginia—are currently using the HCBS CAHPS Survey. States that have used HCBS CAHPS in the past, but are not current users, include Colorado, Georgia, Illinois, Indiana, Kentucky, Maryland, Massachusetts, Minnesota, Mississippi, New Hampshire, New Jersey, New Mexico, New York, Ohio, and Texas.

The American Rescue Plan Act of 2021, Section 9817 spending plans and narratives, identified states who noted funds would be used to adopt HCBS CAHPS or another experience of care survey—the District of Columbia, Hawaii, Indiana, Maine, North Carolina, New Mexico, and Oklahoma. After outreach to these states, Hawaii, Indiana, Maine, and North Carolina identified plans to adopt the HCBS CAHPS Survey. Technical assistance on the HCBS CAHPS Survey has been provided to 27 states and/or plans as of February 2024, including Alabama, California, Colorado, Florida, Georgia, Hawaii, Illinois, Indiana, Iowa, Louisiana, Maine, Massachusetts, Michigan, Missouri, Nevada, New York, North Carolina, Oklahoma, Pennsylvania, Vermont, West Virginia, United Health Care (Arizona, Kansas, and Florida), University of Pittsburgh Medical Center (UPMC) (Pennsylvania), Kepro (West Virginia), AmeriHealth Caritas (Pennsylvania), Simply Healthcare (Florida, and ILS Health (Florida).

*Level of analysis and care setting:* Measures are calculated at the state and managed care program level for services provided in home or community-based care settings.

### **6.2.1 Actions of Measured Entities to Improve Performance**

A second round of HCBS CAHPS Survey administration by Testing Experience and Functional Tools grantees was completed in 2015 with Medicaid HCBS programs serving a variety of adult HCBS populations in Arizona, Colorado, Connecticut, Georgia, Kentucky, Louisiana, Maryland, Minnesota, and New Hampshire. In general, most grantees used the performance data to inform program quality improvement initiatives in the states' Medicaid HCBS waivers. Specific plans for HCBS CAHPS Survey use ranged from comparing participant experience, person-centered care, and other aspects of performance across programs, to identifying quality improvement opportunities, to potentially replacing other surveys used by the state for quality improvement, to exploring whether future managed long-term services and supports programs should use the instrument and resulting performance data. One grantee intended to use the performance data to identify and address improvement opportunities for person-centered care by partnering managed care organizations.

Technical Assistance was provided to state Medicaid agencies around interpreting the results and using the results for quality improvement. Testing Experience and Functional Tools grantees received ongoing technical assistance in a variety of formats (e.g., monthly training webinars, monthly Community of Practice meetings, one-on-one meetings, resource materials posted on a project website). Public users also receive ongoing technical assistance through national training

---

webinars, smaller group presentations, a technical assistance mailbox for making inquiries, materials posted on the CMS webpage for the HCBS CAHPS Survey, and additional resources under development.

The actions measure entities take to improve performance on the measures depend on individual state and various program needs. The following examples of actions to improve performance for measure entities are self-reported by states participating in the early adoption work group meetings hosted by CMS to support Medicaid HCBS and managed long-term services and supports programs with implementing the HCBS CAHPS Survey.

In 2022, Pennsylvania reported identifying areas of improvement for their HCBS program by comparing survey composite scores year over year and against established benchmarks to identify areas of improvement. Pennsylvania analyzed survey data by demographic characteristics to identify disparities in services across their program. All findings of these analyses are presented to managed care plans who develop action plans to address participant concerns. Pennsylvania also used HCBS CAHPS Survey data to compare results between managed care plans to identify areas of concern and establish corrective action plans for specific managed care plans. These survey results and action plans are then presented by managed care plans to stakeholder groups, including executive leadership teams and participant and advocate committees.

In 2022, West Virginia identified survey items for which participant responses fall below an established score threshold and uses these findings to develop plans to improve HCBS based on the survey items, including providing training and education for service providers.

### **6.2.2 Feedback on Measure Performance**

Testing Experience and Functional Tools grantees provided feedback in final reports for field testing conducted from 2013 to 2015. The reports provided information about grantee's experience in fielding the survey, results from data collection and analysis, and how the state is using results.

Planned HCBS CAHPS Survey use ranged from using the survey to assess participant experience, measure provider performance and compliance, reporting quality assurance, and compare quality of services delivered across programs.

Early adoption work group participants have included participants from up to 11 states who have implemented the HCBS CAHPS Survey (i.e., Arizona, Connecticut, Florida, Kentucky, Michigan, New Hampshire, New Jersey, Oklahoma, Pennsylvania, Texas, and West Virginia). During the meetings, states are able to provide feedback on measure performance.

In 2022, Pennsylvania noted that they use survey composite scores over the years against established benchmarks to identify areas of improvement. In 2021, Pennsylvania reports using measure performance to oversee work performance. In 2021, West Virginia reports using survey result performance for quality improvement during provider training with LTSS providers.

As part of early adoption work group meetings, states who are current users can provide best practices for states who are looking to implement. The meetings provide a space for feedback on

implementation.

During a 2021 meeting, current users discussed the process of selecting a survey vendor, which is critical to survey implementation. Pennsylvania shared that they looked for a vendor with experience conducting in-person surveys due to low response rates for surveys conducted in person. West Virginia used agency staff. Connecticut reported using REDCap Portal, a research tool for university data collection. During the meeting, current users provided feedback on implementing a survey vendor. West Virginia shared that they use past survey data to develop provider training which has been helpful for their state. Texas reported using only the phone administration method to improve response rates.

### 6.2.3 Consideration of Measure Feedback

There are no substantive modifications to the measures proposed at this time. The Centers for Medicare & Medicaid Services will continue to elicit feedback from measures users and other stakeholders for potential future modifications.

### 6.2.4 Progress on Improvement

To discuss progress on improvement in HCBS CAHPS Survey use, data from 2022 (Dodson et al., 2023) and 2023 (Dodson et al., 2024) will be referenced. Results collected prior to the 2022 survey year were not included in the measure developer's analyses due to instability in data collection and accuracy during the COVID-19 public health emergency. Progress on improvement on mean performance was seen for 12 measures from 2022 to 2023:

- Global rating of personal care assistant or behavioral health staff (3.7 percent);
- Global rating of homemaker (5.4 percent);
- Recommendation of personal care assistant or behavioral health staff (12.4 percent);
- Recommendation of homemaker (16.9 percent);
- Staff are reliable and helpful (3.2 percent);
- Staff listen and communicate well (3.0 percent);
- Choosing the services that matter to you (1.1 percent)
- Transportation to medical appointments (5.0 percent);
- Personal safety and respect (1.8 percent);
- Planning your time and activities (8.7 percent);
- Sufficient staff to help dress, shower, or bathe (25.9 percent); and
- Sufficient staff to help you with toileting (0.8 percent).

Between 2022 and 2023, the number of accountable entities increased from 17 to 24 (states and managed care programs) and the number of participants for whom data were available increased from 4,731 to 6,053.

#### References:

Dodson, T., Corrothers, M., Yount, N., Sorra, J., Shaller, D. (2023). *The CAHPS® Home and Community-Based Services (HCBS) Survey Database 2023 chartbook*. Agency for Healthcare

---

Research and Quality. <https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/cahps-database/2023-hcbs-chartbook.pdf>

Dodson, T., Corrothers, M., Rubin, J., Vallentine, J., Yount, N., Sorra, J., Shaller, D. (2024). *The CAHPS® Home and Community-Based Services (HCBS) Survey Database 2024 chartbook*. Agency for Healthcare Research and Quality. Chartbook. <https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/cahps-database/2024-hcbs-chartbook.pdf>

### **6.2.5 Unexpected Findings**

To date, no positive or negative findings have been reported by entities using the HCBS CAHPS Survey.

### **7.1 Supplemental Attachment**

[CBE 2967 Supplement.pdf](#)

#### **Developer POC email**

KHammons@mathematica-mpr.com

#### **Measure Developer POC**

Krista Hammons  
Mathematica  
United States

#### **The measure developer is different from the measure steward**

Yes

#### **Steward Address**

Kerry Lida  
7500 Security Boulevard  
Windsor Mill, MD 21244  
United States

#### **Steward Organization**

Centers for Medicare & Medicaid Services

#### **Steward Organization URL**

<https://www.medicaid.gov/medicaid/quality-of-care/quality-of-care-performance-m...>

#### **Steward POC email**

Kerry.Lida@cms.hhs.gov