# Development and Evaluation of Candidate Standardized Patient Assessment Data Elements: Findings from the National Beta Test (Volume 4: Cognitive Function)

**Prepared for**

Charlayne Van
Centers for Medicare & Medicaid Services
Center for Clinical Standards and Quality
Quality Measurement & Health Assessment Group
7500 Security Boulevard
Baltimore, MD 21244

**Prepared by**

Maria Orlando Edelen
Anthony Rodriguez
Sangeeta C. Ahluwalia
Emily K. Chen
Catherine C. Cohen
Sarah Dalton
Jason Michel Etchegaray
Wenjing Huang
Brenda Karkos
David J. Klein
Jaime Madrigano
Monique Martineau
Terry Moore
Patrick Orr
Jessica Phillips
Ben Schalet
Victoria Shier
Susan M. Paddock
Denis Agniel

RR-3004/3-CMS

# Development and Evaluation of Candidate Standardized Patient Assessment Data Elements:

# Findings from the National Beta Test (Volume 4: Cognitive Function)

by

Maria Orlando Edelen
Anthony Rodriguez
Sangeeta C. Ahluwalia
Emily K. Chen
Catherine C. Cohen
Sarah Dalton
Jason Michel Etchegaray
Wenjing Huang
Brenda Karkos
David J. Klein
Jaime Madrigano
Monique Martineau
Terry Moore
Patrick Orr
Jessica Phillips
Ben Schalet
Victoria Shier
Susan M. Paddock
Denis Agniel

Charlayne Van, COR
CMS Contract No. HHSM-500-2013-13014I

For more information on this publication, please visit:
www.rand.org/t/RR3004Z3

This project was sponsored by the Centers for Medicare & Medicaid Services under contract no. HHSM-500-2013-13014I. The research was carried out within the Quality Measurement and Improvement Program in RAND Health Care, a division of the RAND Corporation. For more information, please see www.rand.org/health-care.

# Preface

The Centers for Medicare & Medicaid Services (CMS) contracted with the RAND Corporation to identify and develop standardized patient assessment data elements (SPADEs) for use in the following post-acute care (PAC) patient assessment instruments: the Outcome and Assessment Information Set, used in home health agencies; the Inpatient Rehabilitation Facility Patient Assessment Instrument, used in inpatient rehabilitation facilities; the Long-Term Care Hospital Continuity Assessment Record and Evaluation Data Set, used in long-term care hospitals; and the Minimum Data Set, used in nursing homes and skilled nursing facilities. RAND was tasked with developing and testing data elements within five areas of focus that fall under the clinical categories delineated in the Improving Medicare Post-Acute Care Transformation (IMPACT) Act of 2014: (1) cognitive function and mental status; (2) special services, treatments, and interventions; (3) medical conditions and comorbidities; (4) impairments; and (5) other categories.

This eight-volume report presents background information and results of the National Beta Test, which assessed a set of data elements within the five categories under the IMPACT Act. The National Beta Test was conducted between November 2017 and August 2018. Volume 1 is an executive summary of the material presented in the subsequent volumes. Volume 2 covers the data elements tested; the design; the sampling plan; information on training, recruitment, and retention; information on the data collection process; and the analytic plan. Volume 3 provides a sample description and reports analyses that evaluate the generalizability of results from the National Beta Test sample, both in terms of the representativeness of the facility/agency-level sample to the national population of PAC facilities/agencies, as well as the patients and residents who participated in the National Beta Test relative to the national population of patients and residents receiving PAC in the United States. Volumes 4–8 present the quantitative and qualitative data gathered during testing, as well as interpretations of the results for SPADEs in the following clinical categories: cognitive function (Volume 4), mental status and pain (Volume 5), impairments and special services, treatments, and interventions (Volume 6), and data elements that fall into other clinical categories (care preferences, medication reconciliation, and global health; Volume 7). Volume 8 describes the results and recommendations for SPADEs developed specifically for patients and residents who are unable to communicate (staff assessments of mental status, mood, and pain).

# Contents

# Figures and Tables

## Figures



## Tables

# Abbreviations

| | |
|---|---|
| 6-CIT | Six-Item Cognitive Impairment Test |
| ADLs | activities of daily living |
| AM-PAC | Activity Measure for Post-Acute Care |
| ANOVA | analysis of variance |
| BIMS | Brief Interview for Mental Status |
| CAM | Confusion Assessment Method |
| CARE | Continuity Assessment Record and Evaluation |
| CMS | Centers for Medicare & Medicaid Services |
| CY | calendar year |
| DOTPA | Developing Outpatient Therapy Payment Alternatives |
| EFPT | Executive Function Performance Test |
| FY | fiscal year |
| HHA | home health agency |
| IMPACT | Improving Medicare Post-Acute Care Transformation |
| IRF | inpatient rehabilitation facility |
| IRF-PAI | Inpatient Rehabilitation Facility Patient Assessment Instrument |
| LCDS | Long-Term Care Hospital CARE Data Set |
| LTCH | long-term care hospital |
| MDS | Minimum Data Set |
| MoCA | Montreal Cognitive Assessment |
| Neuro-QoL | Quality of Life in Neurological Disorders |
| OASIS | Outcome and Assessment Information Set |
| PAC | post-acute care |
| PASS | Performance Assessment of Self-Care Skills |
| PC | public comment |
| SD | standard deviation |
| SLUMS | St. Louis University Mental Status |
| SNF | skilled nursing facility |
| SPADE | standardized patient assessment data element |
| TEP | technical expert panel |

# 1. Introduction

The Centers for Medicare & Medicaid Services (CMS) contracted with the RAND Corporation to evaluate candidate standardized patient assessment data elements (SPADEs) in a national field test titled the National Beta Test. The National Beta Test was conducted to evaluate the performance of candidate SPADEs in the clinical categories of (1) cognitive function and mental status; (2) special services, treatments, and interventions; (3) medical conditions and comorbidities; (4) impairments; and (5) other clinical categories, for use in four post-acute care (PAC) settings: home health agencies (HHAs), inpatient rehabilitation facilities (IRFs), long-term care hospitals (LTCHs), and skilled nursing facilities (SNFs).

This is Volume 4 of the final report on the National Beta Test, which includes the identification and testing of candidate SPADEs in the *cognitive function* clinical category. This chapter offers a high-level orientation of the goals, scope, and methods of the National Beta Test. Additionally, this chapter lists the analyses that will be presented for the evaluation of candidate SPADEs in later chapters of this volume.

Candidate SPADEs were identified for the National Beta Test following a series of activities that took place from October 2015 to August 2017, including two Alpha feasibility tests held in select CMS regions,[1] two technical expert panels (TEPs),[2] two subregulatory calls for public comment,[3] and one notice of proposed rulemaking for the Fiscal Year (FY)/Calendar Year (CY) 2018 proposed rules.[4] The results of these activities informed the content and design of the National Beta Test.

The National Beta Test included data collection within 143 PAC facilities/agencies across 14 markets in the United States (listed in Volume 2 of the final report[5]), from November 2017 to August 2018. The overarching goal of the National Beta Test was to evaluate the feasibility, reliability, and validity of candidate SPADEs to identify a subset of data elements for standardization across PAC settings. Candidate SPADEs were considered if they met the requirements of being feasible, clinically useful, and having the potential to improve quality. Trained research nurses and/or staff at participating PAC facilities/agencies administered all National Beta Test assessment protocols. A subset of National Beta Test assessments was

[1] Edelen et al., 2017; Edelen et al., 2018.

[2] RAND Corporation, 2017a; RAND Corporation, 2017b.

[3] CMS, 2016; CMS, 2018.

[4] CMS, 2017a; CMS, 2017b; CMS, 2017c; CMS, 2017d.

[5] Edelen et al., 2019a.

completed by research nurse and facility/agency staff assessor pairs to allow for evaluation of interrater reliability. Other National Beta Test design features allowed for comparison of different look-back time frames for chart review data elements (i.e., on admission [Day 1], and on Days 3, 5, and 7; Discharge Day and Discharge Day minus 2), as well as an evaluation of the assessment of a subset of interview data elements on Days 3, 5, and 7.

To support evaluation of the validity of candidate SPADEs, data collectors documented demographic characteristics of the patient/resident sample (e.g., gender, age). National Beta Test assessment data were merged with CMS routine admission assessment data in the Outcome and Assessment Information Set (OASIS), Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI), Long-Term Care Hospital CARE Data Set (LCDS), and Minimum Data Set (MDS). These assessment data were collected concurrently by the PAC facilities/agencies and submitted to CMS to fulfill PAC regulatory, prospective payment system, and quality reporting program requirements. From these data, a set of variables was selected that reflected the presence of clinical conditions (i.e., sepsis, heart failure, and stroke) and the ability to perform two activities of daily living (ADLs) (toileting [hygiene] and the ability to transfer from lying to sitting [mobility]). These variables, defined in more detail in Volume 3,[6] were selected because they are prevalent, potentially debilitating illnesses or conditions with a high relevance to patients/residents across all four PAC settings. In addition, and crucial for our ability to compare across PAC provider types, these variables were consistently defined across the four PAC settings, although toileting was not available for HHA patients at the time of this study.

Finally, to further support the feasibility and clinical utility of the candidate SPADEs, we solicited the perspectives of research nurses and facility/agency staff assessors on the strengths and weaknesses of collecting the data elements in practice. This feedback was collected as part of the National Beta Test by means of an online survey and focus group discussions.

To evaluate the candidate SPADEs, this report provides the following results and significance tests.

## Feasibility

- Basic descriptive statistics (e.g., frequencies, means, standard deviations [SDs]) for each component, or item, of each data element for all admission data, first combined across settings (overall), and then by setting.
- Extent of missing data for each data element overall. Missing data were minimal and did not vary by setting so they are only briefly summarized.
- Average time to complete the assessment of each data element, for each data element overall and by setting.

[6] Edelen et al., 2019b.

## Reliability

- Interrater reliability for each data element overall and by setting. We examined interrater reliability using a variety of coefficients depending on the response scale of data elements: kappa (dichotomous), weighted kappa (ordinal), and raw percent agreement (all formats).
- For each data element, there are two tables: one reporting kappa and weighted kappa estimates and another reporting raw percent agreement. Interpretation of coefficients follows conventional criteria: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, and 0.81–1.00 is excellent/almost perfect. Because of the impact of prevalence rates on the stability and interpretability of kappa estimates, kappa is not reported for data elements with prevalence rates out of range for stable kappa estimates, as determined by study power calculations. In these cases, kappas are replaced by (—) in the tabulated results.

## Validity

- Frequency tables delineating the association of patient/resident characteristics (i.e., gender, age, length of stay, disposition at discharge), clinical conditions (i.e., sepsis, heart failure, stroke), and two ADLs (i.e., toileting [hygiene] and ability to transfer from lying to sitting [mobility]) with responses to the data element (e.g., Brief Interview for Mental Status [BIMS] categorization). Evaluation of these associations provides a form of construct validity referred to as *known groups validity*, which is demonstrated when a data element can discriminate between two groups in expected ways. Because examination of all data elements by all patient characteristics variables would be prohibitive, we conducted these analyses using data elements representing total scores (e.g., BIMS categorization, Patient Health Questionnaire [PHQ]-9 score, ability to see) where available; when total scores were not available we selected the data element in the set that was both representative and had sufficiently high endorsement rates for significant associations to be observed (e.g., Mechanically Altered Diet). Frequency tables for patients/residents overall are shown in the body of this volume, and setting-level frequencies are contained in the appendix.

## Stability and Change over Time

- Comparison of admission and discharge frequency data for each data element overall and by setting.
- Degree of change in rates or scores depending on the day a patient/resident was assessed within the Day 3, 5, and 7 repeat assessment design. These results are reported for all data elements included in the repeat assessment design overall and by setting.

## Sensitivity to National Representativeness

- Sensitivity analyses for each data element to confirm that performance does not vary according to urbanicity as classified by rural-urban commuting area codes (metropolitan and micropolitan [urban] versus small town and rural [nonurban]), geographic region as defined by the U.S. Census (Northeast, South, Midwest, West), facility ownership (for-profit versus nonprofit), and facility size (above versus below median size for the setting [size analyses not conducted for HHAs]). The results of these sensitivity analyses are included in the appendix. For the most part, differences were not found, and those that were identified are discussed within the specific data element chapter for which a difference emerged.

## Statistical Tests

- Categorical associations were statistically evaluated using chi-square tests of independence and, in the case of ordinal data, Mantel-Haenszel chi-square.[7] Significant results from chi-square tests are reported in the following format: $\left(\chi^2_{(df)} = X.X, p < 0.05\right)$, where *df* are degrees of freedom and the *X*'s are numerical test statistic values. A significant chi-square value (i.e., $p < 0.05$, $p < 0.01$, $p < 0.001$) indicates a significant association between two variables (e.g., age group and BIMS categorization).
- Associations involving continuous variables were statistically evaluated using either an analysis of variance or independent samples t-test to determine whether statistical differences emerged in the continuous variable (e.g., length of stay) as a function of a grouping variable (e.g., BIMS categorization). Significant results from analysis of variance and t-test results are reported in the following formats: $\left(F_{(df)} = X.X, p < 0.001\right)$ or $\left(t_{(df)} = X.X, p < 0.001\right)$, where *df* are degrees of freedom and the *X*'s are numerical test statistic values. When a significant overall effect was found, follow-up independent samples t-tests were often conducted to statistically compare each group value (e.g., to evaluate setting-specific differences in time-to-complete assessments).
- Effect sizes for many of the significant findings are reported using Cohen's *d* to further characterize the importance of statistically significant findings.[8] When reported, a Cohen's *d* value greater than 0.2 was used to indicate a potentially meaningful (i.e., medium to large) effect size.
- When multiple tests were performed (i.e., setting comparisons for time-to-complete assessments, pairwise comparisons between assessment days for repeat assessments, and comparisons between admission to discharge), the probability of finding significant differences by chance increases. To control for this, we calculated

[7] Mantel and Haenszel, 1959.

[8] Cohen, 2013.

corrected significance levels using the Benjamini-Hochberg method, where each significance test is evaluated against an adjusted critical value.[9] We set our desired level of significance at 0.01 to minimize Type I error and increase confidence in significant effects.

[9] Benjamini and Hochberg, 1995.

# 2. Standardized Assessment of Cognitive Function in Post-Acute Care

Cognitive impairment is associated with several disorders, conditions, and injuries (e.g., depression,[10] dementia,[11] stroke,[12] traumatic brain injury[13]). Cognitive impairment has been linked to limitations in the capacity to make informed decisions about health care,[14] lower quality of life,[15] decreased social functioning and ability to maintain personal relationships,[16] and decreased functional status.[17] Cognitive impairment is also associated with an increased likelihood of hospital readmission following discharge to PAC,[18] limitations in the ability to adhere to medication regimens,[19] and increased risk of adverse drug events while receiving home health services.[20] Cognitive impairment can also exacerbate other chronic medical conditions,[21] compromise treatment participation and compliance,[22] slow recovery from injuries and surgeries,[23] and lead to hospitalization or rehospitalization.[24] Because patients/residents in PAC settings are at risk for cognitive impairment, the assessment of cognitive function is important in order to screen for impairment, assess the severity of a disorder, and monitor the progression of symptoms. Understanding an individual's condition and care needs allows for better person-directed care that includes appropriate behavioral or pharmacologic therapies, anticipates the patient's ability to understand and participate in treatments during his or her stay, and identifies appropriate supports at the time of discharge. The timely transmission of information pertaining

[10] Rock et al., 2004.

[11] Hugo and Ganguli, 2014.

[12] Sun, Tan, and Yu, 2014.

[13] Arciniegas, Held, and Wagner, 2002.

[14] Lorig et al., 2001.

[15] Logsdon et al., 2002.

[16] Cruz-Oliver et al., 2012.

[17] Campbell et al., 2005; Heruti et al., 2002; Stuck et al., 1999.

[18] Gage et al., 2012.

[19] Campbell et al., 2012.

[20] Gray, Mahoney, and Blough, 1999.

[21] Feil, Marmon, and Unützer, 2003.

[22] Hayes et al., 2009.

[23] Millar, Asbury, and Murray, 2001; Givens, Sanft, and Marcantonio, 2008.

[24] Buslovich and Kennedy, 2012.

to cognitive function is important for ensuring that the receiving providers and/or the patient, caregiver, or family member has the patient's/resident's information at the time of transfer into the next setting or home. Reliable data elements that assess cognitive function could be used to both initiate and sustain a management program that optimizes a patient's/resident's quality of care and outcomes across the continuum of care.

## Information Gathering

Our review of the literature provided nearly 100 data elements that assessed cognitive function. We prioritized those data elements or sets of data elements that assessed domains of cognition that were most closely related to independent living and daily decisionmaking. For example, we identified several data elements that assessed cognitive ability as related to daily tasks, such as the Performance Assessment of Self-Care Skills (PASS) Medication Management Task and the Developing Outpatient Therapy Payment Alternatives (DOTPA) Continuity Assessment Record and Evaluation (CARE) tool. We considered neurocognitive assessments of executive function and working memory, such as the Montreal Cognitive Assessment (MoCA), the Activity Measure for Post-Acute Care (AM-PAC), Quality of Life in Neurological Disorders (Neuro-QoL), the Executive Function Performance Test (EFPT), and the Assessment of Activities of Daily Living and Instrumental Activities of Daily Living. Additionally, we considered data elements specific to the assessment of executive function, such as a clock drawing task, the St. Louis University Mental Status (SLUMS) exam, the Raven's Progressive Matrices test, and the three-item Mini-Cog assessment.

Several data elements that assess for cognitive function have been used in the four existing PAC assessment instruments. These data elements include the BIMS (included in MDS and IRF-PAI); Medication Management (included in OASIS); Cognitive Functioning, Confusion, and Cognitive/Behavioral/Psychiatric Symptoms (included in OASIS); the Confusion Assessment Method (CAM; included in MDS and LCDS); Expression of Ideas and Wants (included in MDS, OASIS, and IRF-PAI); Behavioral Signs and Symptoms (included in MDS); Staff Assessment of Cognitive Status (included in MDS, LCDS, and IRF-PAI); and Understanding Verbal Content (included in all four existing PAC assessments).

## Stakeholder Feedback and Field Testing

During the focus groups with PAC staff, participants emphasized the impact of cognitive function on successful transitions between care settings and commented on the broad range of instruments to assess cognitive function used across PAC settings and across organizations within the same setting. This feedback aligned with the review of the literature, in which a large number of cognitive assessments with no gold standard in PAC settings was found.

Twelve sets of assessment data elements were presented to the first convening of the TEP for consideration and rating: the BIMS, CAM, Expression of Ideas and Wants, Understanding of

Verbal Content, Observational Assessment of Cognitive Status, MoCA, Mini-Cog, SLUMS, the Six-Item Cognitive Impairment Test (6-CIT), the General Practitioner Assessment of Cognition, the PASS Medication Management Task, and EFPT.[25] During a rating exercise, TEP members independently evaluated each data element across five dimensions (potential for improving quality, validity, reliability, feasibility for use in PAC, and utility for describing case mix) by assigning numeric ratings on a five-point scale (1 = poor, 5 = excellent). The TEP provided the highest level of overall support to the BIMS, perceiving it to be "very good" across all rating dimensions, but also gave high ratings to the MoCA for its validity, reliability, potential to improve quality, and utility for describing case mix, though there were concerns about the length of the assessment. The TEP also rated the CAM highly on its potential for improving quality, validity, and reliability, and provided moderate support for the DOTPA CARE, which was not initially included in the rating exercise but was raised in the discussion and subsequently rated during the second convening of the TEP. At that meeting, the DOTPA received an overall score of 3.1, indicating that it was somewhat but not highly supported. Overall, the TEP recommended a screening approach to identify persons for longer assessment—that is, a brief assessment of cognitive function that would provide a general indication of cognitive issues that would then lead to a more comprehensive assessment. The TEP agreed that a standardized cognitive assessment would serve as the beginning of a more comprehensive cognitive assessment for PAC patients/residents.

Five cognitive function data elements were posted for public comment in 2016: the BIMS, the CAM, Expression of Ideas and Wants, Understanding Verbal Content, and Behavioral Signs and Symptoms. Each of these data elements has strong evidence of feasibility and validity for cross-setting use, and commenters expressed specific support for each of them. In general, commenters affirmed the importance of appropriate cognitive assessment. Three additional cognitive function data elements (DOTPA CARE, Complex Sentence Repetition, and the PASS Medication Management Task) were presented for public comment in 2017. Several commenters provided support for DOTPA CARE and the PASS Medication Management Task for helping to identify patients who could have difficulties implementing strategies related to medication management, maintenance of chronic conditions, and other aspects pertaining to self-care and safety when in the home. These patients could require services to remediate or compensate for impairments. However, commenters raised concerns regarding burden and feasibility. Commenters also provided support for Complex Sentence Repetition, but there were concerns related to its applicability with different types of patients, such as those with neurological diagnoses, children, and nonverbal patients/residents.

The first Alpha feasibility test (Alpha 1) contained three executive function tasks (Trail Making, Serial 7's, and Complex Sentence Repetition). Although cross-setting feasibility data

[25] RAND Corporation, 2017a.

were already available for the BIMS, the BIMS was included in Alpha 1 specifically to investigate the benefit of assessing executive functions in addition to the content currently assessed with the BIMS. The Alpha 1 test also allowed the ability to test whether a skip pattern is appropriate after administration of the BIMS for later data collection. All data elements performed well and were found to be feasible across PAC settings.

The second Alpha feasibility test (Alpha 2) evaluated the PASS Medication Management Task and DOTPA CARE. Assessors reported that both instruments were straightforward to administer but burdensome, with each taking a relatively considerable amount of time (five to seven minutes) to complete. Although interrater reliability was high overall for the PASS Medication Management Task, assessors reported that many patients in HHA and IRF settings required assistance to complete it, and missing data were prevalent. Feedback from assessors indicated that DOTPA CARE might not be as relevant to patients/residents of LTCHs and SNFs, and interrater reliability was highly variable among those data elements, ranging from very low to excellent.

## Candidate SPADEs in the National Beta Test

Taking together findings from the literature, input from the TEP and other subject-matter experts, and considerations of feasibility and burden for implementation, four data elements were selected that assessed aspects of cognitive function and were tested in the National Beta Test. These data elements are shown in Table 2.1. This table also lists the evaluative input opportunities for each data element for this study, specific design features relevant to the data element for the National Beta Test, and indicates the data element's current use in any of the four PAC assessments (OASIS, IRF-PAI, LCDS, and MDS).

**Table 2.1. Cognitive Function Data Elements Evaluated in the National Beta Test Communicative Sample**

| Data Element | Input Opportunities | National Beta Test Inclusion Notes | Current Assessment Instrument Use |
|---|---|---|---|
| BIMS | Public Comment (PC) 1, FY/CY 2018 proposed rule | • Included in Day 3-5-7 test | • IRF-PAI<br>• MDS |
| CAM | PC 1, FY/CY 2018 proposed rule | • Included in Day 3-5-7 test | • LCDS<br>• MDS |
| Expression and Understanding | PC 1 | • Two versions tested[a]<br>• Included in Day 3-5-7 test | • OASIS[b]<br>• IRF-PAI<br>• LCDS<br>• MDS |
| Behavioral Signs and Symptoms | PC 1, FY/CY 2018 proposed rule; Alpha 2, PC 2 | • Included in Day 3-5-7 test | • MDS |

NOTE: Assessment of these data elements in the National Beta Test was limited to communicative patients/residents (defined as those who could make themselves understood by any means; see Volume 2 for more detail).

[a] Two versions of the Expressions and Understanding data elements were evaluated in different samples. The three–data element version, which consists of Speech Clarity, Makes Self Understood, and Ability to Understand Others, is identical to the data elements that are currently used in the MDS. The two–data element version, which consists of Expression of Ideas and Wants and Understanding Verbal Content, is identical to the data elements that are currently used in the IRF-PAI and LCDS.

[b] Item M1230 removed from the OASIS-D, effective January 1, 2019.

# 3. Brief Interview for Mental Status

## Data Element Description

The BIMS is a performance-based cognitive assessment that assesses repetition, recall with and without prompting, and temporal orientation. As described in Chapter 2, cognitive impairment has been linked to limitations in the capacity to make informed decisions about health care[26] and adhere to medication regimens, [27] a lower quality of life,[28] decreased social functioning, decreased ability to maintain personal relationships,[29] and decreased functional status.[30] Conducting cognitive assessments is critically important to screen for cognitive impairment, rate severity of disorder, develop a care plan, and monitor progression.

The BIMS is completed via a patient/resident interview to determine how the patient/resident performs on a series of tasks. The BIMS was developed as a brief, simply scored, structured screen of cognitive function as part of the major revision of the MDS from version 2.0 to version 3.0.[31] The BIMS is currently used in the MDS 3.0 and in the IRF-PAI. Observation of patient/resident behavior during the BIMS or other similar assessments of cognitive function inform the assessment of the CAM. In the National Beta Test, the BIMS was included in the repeat assessment evaluation and therefore was evaluated repeatedly on the same patient/resident by the same assessor on Days 3, 5, and 7, per the repeat assessment design. The BIMS data element is shown in Figure 3.1.

[26] Lorig et al., 2001.

[27] Campbell et al., 2012.

[28] Logsdon et al., 2002.

[29] Cruz-Oliver et al., 2012.

[30] Campbell et al., 2005; Heruti et al., 2002; Stuck et al., 1999.

[31] Saliba et al., 2012; Chodosh et al., 2008.

**Figure 3.1. Brief Interview for Mental Status**

**B1a. Repetition of Three Words**

**ASK PATIENT/RESIDENT:** "I am going to say three words for you to remember. Please repeat the words after I have said all three. The words are: sock, blue and bed. Now tell me the three words."

Number of words repeated by patient/resident after first attempt:

- ☐ 3 = Three
- ☐ 2 = Two
- ☐ 1 = One
- ☐ 0 = None or no answer

**AFTER THE PATIENT'S/RESIDENT'S FIRST ATTEMPT SAY:** "I will repeat each of the three words with a cue and ask you about them later: sock, something to wear; blue, a color; bed, a piece of furniture." **YOU MAY REPEAT THE WORDS UP TO TWO MORE TIMES.**

**B1b. Year, Month, Day**

**ASK PATIENT/RESIDENT:** "Please tell me what year it is right now."

Patient's/Resident's answer is:

- ☐ 3 = Correct
- ☐ 2 = Missed by 1 year
- ☐ 1 = Missed by 2 to 5 years
- ☐ 0 = Missed by more than 5 years or no answer

**B1c. ASK PATIENT/RESIDENT:** "What month are we in right now?"

Patient's/Resident's answer is:

- ☐ 2 = Accurate within 5 days
- ☐ 1 = Missed by 6 days to 1 month
- ☐ 0 = Missed by more than 1 month or no answer

**B1d. ASK PATIENT/RESIDENT:** "What day of the week is today?"

Patient's/Resident's answer is:

- ☐ 1 = Accurate
- ☐ 0 = Incorrect or no answer

| |
|---|
| **ASK PATIENT/RESIDENT:** "Let's go back to the first question. What were those three words that I asked you to repeat?" **IF UNABLE TO REMEMBER A WORD, GIVE CUE (I.E., SOMETHING TO WEAR; A COLOR; A PIECE OF FURNITURE) FOR THAT WORD.**<br><br>**B1e.** Recalls "sock?"<br><br>☐ 2 = Yes, no cue required<br>☐ 1 = Yes, after cueing ("something to wear")<br>☐ 0 = No, could not recall or no answer |
| **B1f.** Recalls "blue?"<br><br>☐ 2 = Yes, no cue required<br>☐ 1 = Yes, after cueing ("a color")<br>☐ 0 = No, could not recall or no answer |
| **B1g.** Recalls "bed?"<br><br>☐ 2 = Yes, no cue required<br>☐ 1 = Yes, after cueing ("a piece of furniture")<br>☐ 0 = No, could not recall or no answer |

## Testing Objectives

Basic descriptive statistics (e.g., frequencies, means, SDs) are presented for BIMS admission data to characterize the rates of impairment for patients/residents in each setting, as well as for the overall sample. To examine known groups validity, we used the BIMS impairment category (Intact, Moderately Impaired, Severely Impaired) to characterize cognitive impairment at admission by patient/resident characteristics and clinical groups of interest. For admission data, feasibility (frequencies, rates of missingness and time to complete) and interrater reliability (kappa and percent agreement) were examined. The BIMS was also administered repeatedly to a subset of the sample on the same patient/resident by the same assessor per the repeat assessment design, which tests the effects of conducting the assessment three, five, or seven days following admission. As such, an additional objective was to understand whether there were significant and meaningful differences in rates or scores depending on the day a patient/resident was assessed. Lastly, frequencies at admission and discharge were compared to inform stability or possible change over time.

# Results

## *Feasibility*

### Frequencies/Missing

Table 3.1 shows the percentage of responses at admission for each BIMS data element overall and by setting. The BIMS was administered to 3,062 of the 3,121 patients/residents, or 98 percent of the admission sample: 646 in HHAs, 786 in IRFs, 496 in LTCHs, and 1,134 in SNFs. Among those who were administered the BIMS, missing data at the data element level ranged from 0.3 to 1.7 percent overall, with minimal setting differences. Results for the BIMS show that, overall, 76 percent of patients/residents had intact (i.e., not impaired) cognition, while 18 percent were moderately impaired and 5 percent were severely impaired. Results also indicate that there was a significant association between setting type and BIMS categorization $\left(\chi^2_{(6)} = 39.19, p < 0.001\right)$: SNFs and LTCHs had more cognitively impaired patients/residents compared with IRFs and HHAs.

**Table 3.1. Overall and Setting-Specific Response Frequencies for BIMS Data Elements at Admission (percent)**

| Data Element | HHA (*n* = 646) | IRF (*n* = 786) | LTCH (*n* = 496) | SNF (*n* = 1,134) | Overall (*n* = 3,062) |
|---|---|---|---|---|---|
| Number of words repeated after first attempt (b1a) | | | | | |
| Three | 94 | 96 | 91 | 94 | 94 |
| Two | 4 | 3 | 4 | 4 | 4 |
| One | 1 | 1 | 2 | 1 | 1 |
| None or no answer | 0 | 1 | 3 | 1 | 1 |
| Recalls current year (b1b) | | | | | |
| Correct | 89 | 94 | 88 | 87 | 89 |
| Missed by 1 year | 2 | 1 | 4 | 2 | 2 |
| Missed by 2–5 years | 1 | 1 | 1 | 2 | 1 |
| Missed by more than 5 years or no answer | 7 | 4 | 8 | 9 | 7 |
| Recalls current month (b1c) | | | | | |
| Accurate within 5 days | 94 | 93 | 90 | 90 | 91 |
| Missed by 6 days to 1 month | 3 | 3 | 2 | 4 | 3 |
| Missed by more than 1 month or no answer | 4 | 4 | 8 | 6 | 5 |
| Recalls current day of week (b1d) | | | | | |
| Accurate | 88 | 84 | 77 | 76 | 81 |
| Incorrect or no answer | 12 | 16 | 23 | 24 | 19 |

| Data Element | HHA (*n* = 646) | IRF (*n* = 786) | LTCH (*n* = 496) | SNF (*n* = 1,134) | Overall (*n* = 3,062) |
|---|---|---|---|---|---|
| Recalls "sock" (b1e) | | | | | |
| Yes, no cue required | 80 | 84 | 78 | 76 | 79 |
| Yes, after cue | 9 | 5 | 9 | 9 | 8 |
| No recall or answer | 11 | 11 | 13 | 15 | 13 |
| Recalls "blue" (b1f) | | | | | |
| Yes, no cue required | 84 | 85 | 78 | 79 | 81 |
| Yes, after cue | 11 | 11 | 12 | 13 | 12 |
| No recall or answer | 6 | 5 | 10 | 8 | 7 |
| Recalls "bed" (b1g) | | | | | |
| Yes, no cue required | 73 | 75 | 64 | 66 | 70 |
| Yes, after cue | 12 | 10 | 12 | 14 | 12 |
| No recall or answer | 14 | 14 | 24 | 19 | 18 |
| BIMS Composite Score, Mean (SD) | 13.5 (2.4) | 13.7 (2.2) | 12.9 (3.1) | 12.9 (2.9) | 13.3 (2.7) |
| BIMS Impairment Category (based on responses to b1a–b1g)[a] | | | | | |
| Intact | 80 | 82 | 73 | 72 | 76 |
| Moderately impaired | 17 | 15 | 19 | 22 | 18 |
| Severely impaired | 4 | 3 | 7 | 7 | 5 |

[a] Significant ($p < 0.01$) association between setting type and BIMS Impairment Category, as determined by chi-square test of independence.

### Known Groups Validity

Comparing the performance of patients/residents on the BIMS with other patient/resident characteristics adds information about the validity of the BIMS data elements. If known or logical associations between patients/resident characteristics and data elements are observed in data from the National Beta Test, this observation contributes to the evidence that the data elements are valid or assessing the construct that they are intended to capture.

Table 3.2 shows BIMS impairment categories (Intact, Moderately Impaired, Severely Impaired) for the overall admission sample, stratified by patient/resident characteristics and clinical groups as described in Chapter 1: gender (male or female, as documented by National Beta Test assessor), age (as categorized into the following ranges: 18–44, 45–64, 65–74, 75–89, 90 and over), length of stay (in days), disposition at discharge (e.g., to another PAC setting, home, to hospital), sepsis, heart failure, stroke, and two ADLs—toileting (not available for HHA patients) and ability to transfer from lying to sitting. As a reminder, these clinical conditions were chosen based on their common occurrence across settings, their frequent relationship with many of the data elements tested in the National Beta Test, and their availability in all four settings (i.e., equivalent information was collected on the OASIS, IRF-PAI, LCDS, and MDS). Setting-specific results are presented in the appendix (Tables A.1–A.4).

Based on the research literature, we generated several hypotheses or expectations for associations between the data elements and the patient/resident characteristics. We expected BIMS impairment categories to be related to age, heart failure, stroke, needing assistance with toileting, and ability to transfer from lying to sitting, such that more-impaired patients/residents would tend to be older,[32] be more likely to have heart failure[33] or have had a stroke,[34] and have less independence in ADLs.[35]

Across the full sample, we observed significant associations for all patient/resident characteristics and clinical groups with BIMS impairment categorization, except for sepsis and heart failure. This is consistent with prior research, which observes that individuals with higher levels of cognitive impairment are older, sicker, and more dependent with ADLs than individuals with little or no cognitive impairment. We review the statistical associations between variables next.

**Table 3.2. Frequencies for BIMS Impairment Categorization by Patient/Resident Characteristics and Clinical Groups (percent)**

| Patient/Resident Characteristics and Clinical Groups | Intact | Moderately Impaired | Severely Impaired |
|---|---|---|---|
| Gender (*n* = 2,866[a]) | | | |
| Male (*n* = 1,175) | 73.1 | 21.1 | 5.8 |
| Female (*n* = 1,691) | 79.5 | 16.7 | 4.7 |
| Age (*n* = 2,856[a]) | | | |
| 18–44 (*n* =38) | 92.1 | 7.9 | 0.0 |
| 45–64 (*n* = 306) | 84.6 | 13.7 | 1.6 |
| 65–74 (*n* = 896) | 82.6 | 14.6 | 2.8 |
| 75–89 (*n* = 1,311) | 71.9 | 21.4 | 6.7 |
| 90 or older (*n* = 305) | 66.2 | 23.9 | 9.8 |
| Length of stay (*n* = 2,529[a]; mean, SD) | 21.2 (12.6) | 22.1 (12.9) | 25.0 (13.2) |

[32] Unverzagt et al., 2001.

[33] Vogels et al., 2007.

[34] Sun, Tan, and Yu, 2014.

[35] Carpenter et al., 2006.

| Patient/Resident Characteristics and Clinical Groups | Intact | Moderately Impaired | Severely Impaired |
|---|---|---|---|
| Disposition at discharge (*n* = 2,818[a]) | | | |
| Home (*n* = 1,311) | 80.2 | 15.9 | 3.8 |
| Hospital (*n* = 197) | 69.5 | 22.8 | 7.6 |
| Hospice (*n* = 38) | 71.1 | 23.7 | 5.3 |
| SNF (*n* = 273) | 70.7 | 22.0 | 7.3 |
| IRF (*n* = 49) | 73.5 | 16.3 | 10.2 |
| HHA (*n* = 612) | 79.6 | 17.5 | 2.9 |
| LTCH (*n* = 13) | 38.5 | 46.2 | 15.4 |
| Other (*n* = 325) | 67.7 | 22.2 | 10.2 |
| Clinical conditions (*n* = 2,215) | | | |
| Sepsis | | | |
| Yes (*n* = 150) | 72.7 | 18.7 | 8.7 |
| No (*n* = 2,065) | 76.7 | 18.6 | 4.8 |
| Heart failure | | | |
| Yes (*n* = 382) | 73.6 | 21.5 | 5.0 |
| No (*n* = 1,833) | 77.0 | 18.0 | 5.1 |
| Stroke | | | |
| Yes (*n* = 200[a]) | 64.0 | 24.5 | 11.5 |
| No (*n* = 2,015) | 77.6 | 18.0 | 4.4 |
| Hygiene—Toileting (*n* = 1,505[a])[b] | | | |
| Independent (*n* = 69) | 87.0 | 13.0 | 0.0 |
| Setup or clean-up assistance (*n* = 78) | 87.2 | 12.8 | 0.0 |
| Supervision or touching assistance (*n* = 317) | 80.4 | 15.8 | 3.8 |
| Partial/moderate assistance (*n* = 360) | 81.1 | 14.4 | 4.4 |
| Substantial/maximal assistance (*n* = 330) | 72.4 | 21.5 | 6.1 |
| Dependent (*n* = 351) | 70.1 | 21.7 | 8.3 |
| Mobility—Transfer from lying to sitting (*n* = 1,846[a]) | | | |
| Independent (*n* = 186) | 85.0 | 13.4 | 1.6 |
| Setup or clean-up assistance (*n* = 112) | 81.3 | 16.1 | 2.7 |
| Supervision or touching assistance (*n* = 523) | 80.1 | 17.2 | 2.7 |
| Partial/moderate assistance (*n* = 601) | 78.9 | 17.1 | 4.0 |
| Substantial/maximal assistance (*n* = 292) | 74.0 | 19.2 | 6.9 |
| Dependent (*n* = 132) | 63.6 | 22.7 | 13.6 |

[a] Significant ($p < 0.05$) associations with BIMS impairment category as indicated by chi-square tests of independence (analysis of variance [ANOVA] for length of stay).
[b] Toileting data not available for HHA patients

*Gender and Age*

- Cognitive impairment, overall, was significantly associated with gender $\left(\chi^2_{(2)} = 11.32, p < 0.001\right)$, with males tending to have more impairment than females: 79 percent of females were categorized as having intact cognition, compared with 73 percent of males; 17 percent of females were categorized as moderately impaired, compared with 21 percent of males; and 5 percent of females were categorized as severely impaired, compared with 6 percent of males. Similar trends of males having higher impairment than females were observed at the setting level for all settings; however, this trend was only significant at the setting level in the IRF setting $\left(\chi^2_{(2)} = 13.22, p < 0.01\right)$. *We did not expect to find this association, which is difficult to interpret without controlling for age and socioeconomic status.*
- Age was associated with cognitive impairment, overall $\left(\chi^2_{(8)} = 77.98, p < 0.001\right)$, which tended to increase with increasing age, aligning with expectations. The largest proportions of patients/residents who were moderately or severely impaired were in the 75–89 and 90-plus age groups. At the setting level, cognitive impairment was also positively and significantly associated with age among patients/residents in all four settings (HHA: $\chi^2_{(8)} = 32.79, p < 0.001$; IRF: $\chi^2_{(8)} = 16.43, p < 0.05$; LTCH: $\chi^2_{(8)} = 26.46, p < 0.001$; SNF: $\chi^2_{(8)} = 29.57, p < 0.001$). As in the overall results, impairment rates at the setting level tended to increase with increasing age. *This finding is consistent with our expectations that older patients/residents would show higher rates of cognitive impairment and speaks to the valid performance of the BIMS.*

*Length of Stay and Disposition at Discharge*

- Length of stay, overall, was significantly associated with impairment severity $\left(F_{(2,2526)} = 5.52, p < 0.01\right)$: Severely impaired patients/residents, on average, had a longer length of stay (mean = 25 days, SD = 13.16) than moderately impaired patients/residents (mean = 22 days, SD = 12.9; $t_{(2526)} = 2.2, p < 0.05$) and intact patients/residents (mean = 21.2 days, SD = 12.6; $t_{(2526)} = 3.14, p < 0.01$). Length of stay did not statistically differ between moderately impaired and intact patients/residents. At the setting level, length of stay was significantly associated with cognitive impairment in IRFs $\left(F_{(2,713)} = 3.75, p < 0.05\right)$ and SNFs $\left(F_{(2,929)} = 4.61, p < 0.01\right)$, in which severely impaired patients had a longer length of stay than intact patients. *We did not expect to find this association, but it is consistent with the broader finding of patients/residents with greater impairment needing more support and, therefore, more time receiving PAC services.*
- Overall, disposition at discharge was associated with cognitive impairment $\left(\chi^2_{(14)} = 64.93, p < 0.001\right)$. Relative to patients/residents discharged to all other placements, those discharged to LTCHs were more likely to be classified as either moderately or severely impaired$\left(\chi^2_{(14)} = 32.41, p < 0.01\right)$. This significant pattern was also observed in IRFs $\left(\chi^2_{(14)} = 55.52, p < 0.001\right)$ and SNFs $\left(\chi^2_{(14)} = 47.85, p < 0.001\right)$. However, disposition at discharge was not associated with cognitive impairment in HHA patients. *We did not expect to find this association,*

*but it is plausible that patients/residents with greater impairment would also have higher levels of medical needs (e.g., subsequent to stroke or traumatic brain injury) that would require the types of medical supports offered by an LTCH.*

*Clinical Conditions*

- Overall, there were no associations between sepsis and BIMS categorization. Specifically, 27 percent of those with sepsis and 24 percent without were either moderately or severely impaired. Similarly, there were no associations between sepsis and BIMS categorization in each of the four settings. *This finding is consistent with our expectation of no relationship between the BIMS and this clinical condition.*
- Overall, there were also no associations between heart failure and BIMS categorization. Specifically, 26 percent of those with heart failure and 23 percent of those without were either moderately or severely impaired. Similarly, there were no associations between heart failure and BIMS categorization in each of the four settings. *This finding is consistent with our expectation of no relationship between the BIMS and this clinical condition.*
- There was a significant overall association between stroke and BIMS impairment categorization $\left(\chi^2_{(2)} = 26.66, p < 0.001\right)$: 37 percent of those with stroke were either moderately or severely impaired, compared with 23 percent of nonstroke patients/residents. This association was also evident in a similar pattern among patients/residents in IRF $\left(\chi^2_{(2)} = 6.58, p < 0.05\right)$, LTCH $\left(\chi^2_{(2)} = 34.94, p < 0.001\right)$, and SNF $\left(\chi^2_{(2)} = 11.43, p < 0.01\right)$ settings, but not in HHAs. *This finding is consistent with our expectation and supports the validity of the BIMS to capture the types of cognitive impairment that are sometimes the consequence of stroke.*

*ADLs: Toileting and Ability to Transfer from Lying to Sitting*

- BIMS impairment categorization was also associated (as expected) with independence levels on both toileting $\left(\chi^2_{(10)} = 33.04, p < 0.01\right)$ and ability to transfer from lying to sitting $\left(\chi^2_{(10)} = 47.25, p < 0.001\right)$, such that rates of intact cognition according to the BIMS tended to increase as patients'/residents' level of independence on these ADLs increased. For example, rates of intact cognition were 87 percent and 85 percent among patients/residents rated as independent on toileting and ability to transfer from lying to sitting, respectively, whereas rates for intact cognition were 70 percent and 64 percent on toileting and ability to transfer from lying to sitting, respectively, among those rated as completely dependent on these two ADLs. This trend was also observed among SNF residents for toileting $\left(\chi^2_{(10)} = 19.90, p < 0.05\right)$ and among LTCH patients for both toileting $\left(\chi^2_{(10)} = 23.79, p < 0.01\right)$ and ability to transfer from lying to sitting $\left(\chi^2_{(10)} = 28.43, p < 0.01\right)$. *This finding is consistent with our expectations and supports the validity of the BIMS to capture cognitive impairment that relates to independence in ADLs.*

### Time to Complete

Table 3.3 shows the average time to complete the BIMS overall and by setting. On average, the entire BIMS took 2.2 minutes (SD = 1.2) to complete. Setting-specific times to complete ranged from 1.8 minutes (SD = 0.9) in IRFs to 2.4 minutes (SD = 1.2) in HHAs. Time to complete was associated with setting type $\left(F_{(3,1804)} = 23.17, p < 0.001\right)$: The BIMS took significantly less time to complete in IRFs than in the other three settings—HHA $\left(t_{(1804)} = 7.50, p < 0.001\right)$, LTCH $\left(t_{(1804)} = 6.27, p < 0.001\right)$, and SNF $\left(t_{(1804)} = 5.32, p < 0.001\right)$. There were no significant differences in time to complete the BIMS among HHAs, SNFs, and LTCHs.

**Table 3.3. Time to Complete the BIMS (minutes)**

| Characteristic | HHA (*n* = 445) | IRF (*n* = 537) | LTCH (*n* = 332) | SNF (*n* = 494) | Overall (*n* = 1,808) |
|---|---|---|---|---|---|
| Mean time to complete (SD) | 2.4 (1.2) | 1.8 (0.9)[a] | 2.3 (1.2) | 2.2 (1.1) | 2.2 (1.2) |

[a] Significantly lower time to complete in IRFs than in any other setting.

Time to complete was also evaluated according to urbanicity (urban versus nonurban), geographic region (Northeast, South, Midwest, West), facility ownership (for-profit versus nonprofit), and facility size (above or below setting-type median) to evaluate the generalizability of these performance results (see Tables A.5–A.8 in the appendix). Although assessors in the Midwest took significantly longer to complete the BIMS relative to the West and Northeast, the effect sizes were quite small (0.18 and 0.13). No other significant differences were found for time to complete the BIMS in these sensitivity analyses.

## *Interrater Reliability*

Table 3.4 shows kappa interrater reliability coefficients for the BIMS overall and by setting. As described in more detail in Volume 3, paired assessment data for interrater reliability evaluation were collected on a subset of the National Beta Test admission sample of patients/residents according to setting-level target totals. For example, each participating LTCH was asked to conduct 20 paired assessments to contribute to interrater reliability. Inclusion in interrater reliability data collection depended on paired facility staff and research nurse assessors' ability to schedule assessments. Kappas were computed on the 966 patients/residents who were assessed by research nurse and facility/agency staff assessor pairs: 199 in HHAs, 259 in IRFs, 238 in LTCHs, and 270 in SNFs. Overall, kappas for the BIMS tended to be excellent, ranging from 0.83 to 0.93, with minimal setting differences. With the exception of one data element (Recalls "blue," b1f), which had a kappa value of 0.78 in SNFs, all kappas across settings were excellent. Kappas ranged from 0.84 to 0.94 in HHAs, 0.81 to 0.91 in IRFs, 0.87 to 0.93 in LTCHs, and 0.78 to 0.93 in SNFs.

Interrater reliability (kappa) was also evaluated according to urbanicity (urban versus nonurban), geographic region (Northeast, South, Midwest, West), facility ownership (for-profit versus nonprofit), and facility size (above or below setting-type median) to evaluate the generalizability of these performance results (see Tables A.9–A.12 in the appendix). No noteworthy differences were found for interrater reliability of the BIMS in these sensitivity analyses.

**Table 3.4. Interrater Reliability Kappa or Weighted Kappa for BIMS Data Elements**

| Data Element | HHA (*n* = 199) | IRF (*n* = 259) | LTCH (*n* = 238) | SNF (*n* = 270) | Overall (*n* = 966) |
|---|---|---|---|---|---|
| Number of words repeated after first attempt (b1a) | — | — | — | — | — |
| Recalls current year (b1b) | 0.88 | — | 0.90 | 0.93 | 0.90 |
| Recalls current month (b1c) | — | — | 0.89 | 0.86 | — |
| Recalls current day of week (b1d) | 0.92 | 0.81 | 0.91 | 0.86 | 0.88 |
| Recalls “sock” (b1e) | 0.87 | 0.91 | 0.91 | 0.91 | 0.91 |
| Recalls “blue” (b1f) | 0.84 | 0.82 | 0.87 | 0.78 | 0.83 |
| Recalls “bed” (b1g) | 0.93 | 0.90 | 0.93 | 0.93 | 0.93 |
| BIMS impairment category (based on responses to b1a–b1g) | 0.94 | 0.85 | 0.91 | 0.91 | 0.91 |

NOTES: Interrater reliability not shown for data elements with proportions out of range for stable kappa estimate (per study power calculations). Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

Table 3.5 shows percent agreement for the BIMS data elements overall and by setting. Overall percent agreement was high for all data elements, ranging from 94 percent to 98 percent, with minimal setting differences. At the setting level, percent agreement ranged from 94 percent to 98 percent in HHAs, 94 percent to 99 percent in IRFs, 93 percent to 97 percent in LTCHs, and 91 percent to 97 percent in SNFs. Once again, the lowest overall agreement (94 percent) was for the data element asking about recollection of the word “blue.”

**Table 3.5. Interrater Reliability—Percent Agreement for BIMS Data Elements**

| Data Element | HHA (*n* = 199) | IRF (*n* = 259) | LTCH (*n* = 238) | SNF (*n* = 270) | Overall (*n* = 966) |
|---|---|---|---|---|---|
| Number of words repeated after first attempt (b1a) | 96 | 97 | 96 | 96 | 96 |
| Recalls current year (b1b) | 97 | 98 | 97 | 97 | 98 |
| Recalls current month (b1c) | 98 | 99 | 97 | 96 | 98 |
| Recalls current day of week (b1d) | 98 | 94 | 97 | 95 | 96 |
| Recalls "sock" (b1e) | 94 | 97 | 95 | 96 | 95 |
| Recalls "blue" (b1f) | 95 | 95 | 93 | 91 | 94 |
| Recalls "bed" (b1g) | 96 | 95 | 95 | 96 | 96 |
| BIMS impairment category (based on responses to b1a–b1g) | 97 | 95 | 95 | 95 | 96 |

### *Day 3, 5, and 7 Repeat Assessment Evaluation*

Table 3.6 summarizes patterns of change across the repeat assessment days. Patterns are characterized as "no change" (scores stay the same across assessment days), "improve" (steady improvement across assessment days), "worsen" (steady decline over assessment days), and "fluctuate" (scores go up and down across assessment days). As described in more detail in Volume 3, repeat assessment data were collected on a subset of the National Beta Test admission sample of patients/residents according to setting-level target totals. For example, each participating HHA was asked to contribute five patients for repeat assessment. Inclusion in repeat assessment data collection depended on assessor ability to complete the initial assessment on Day 3, availability of the assessor to return for repeat assessments on Days 5 and 7, and willingness of patient/resident to complete multiple assessments. For the BIMS, 480 patients/residents completed repeat assessments at Days 3, 5, and 7 of admission: 95 in HHAs, 122 in IRFs, 71 in LTCHs, and 192 in SNFs. Responses to BIMS data elements were very similar regardless of the day on which the BIMS was administered. Overall, between 65 percent and 92 percent of answers stayed the same across the three assessment points (see "no change" rows in table). Statistically, there were no significant differences across assessment days for "number of words repeated after first attempt" and ability to recall "blue" and the current year, month, and day of the week.

Differences were seen, however, in the overall sample in rates of recall of "sock" and "bed," which tended to improve over time (15 percent and 17 percent, respectively). The effect of assessment day on recollection of the word "sock" was significant $\left(F_{(2,590)} = 11.94, p < 0.001\right)$: Compared with Day 3, patients/residents were 1.45 times more likely on Day 5 $\left(t_{(591)} = 2.96, p < 0.003\right)$ and 1.97 times more likely on Day 7 $\left(t_{(591)} = 4.81, p < 0.001\right)$ to recall the word "sock." Similarly, there was a significant effect of assessment day on recollection of the word "bed" $\left(F_{(2,590)} = 14.22, p < 0.001\right)$: Compared with Day 3, patients/residents were

1.43 times more likely on Day 5 $\left(t_{(591)} = 3.36, p < 0.001\right)$ and 1.87 times more likely on Day 7 $\left(t_{(591)} = 5.19, p < 0.001\right)$ to recall the word "bed." These differences in recall resulted in a significant effect of assessment day on overall BIMS score (mean scores not shown in table; $F_{(2,591)} = 13.84, p < 0.001$): Compared with Day 3, there was significant improvement in overall BIMS score on Day 5 $\left(t_{(591)} = 3.73, p < 0.001\right)$ and Day 7 $\left(t_{(591)} = 5.06, p < 0.001\right)$. Similarly, for BIMS categorization (e.g., intact versus moderate or severe impairment), there was a significant effect of assessment day $\left(F_{(2,590)} = 11.28, p < 0.001\right)$: Patient/resident categorization improved on Day 5 $\left(t_{(591)} = 3.29, p < 0.001\right)$ and Day 7 $\left(t_{(591)} = 4.60, p < 0.001\right)$, compared with Day 3. For the remaining data elements, there were no significant effects of assessment day.

**Table 3.6. Day 3, 5, and 7 Repeat Assessment Results for BIMS Data Elements (percent)**

| Data Element | HHA (*n* = 95) | IRF (*n* = 122) | LTCH (*n* = 71) | SNF (*n* = 192) | Overall (*n* = 480) |
|---|---|---|---|---|---|
| Number of words repeated after first attempt (b1a) | | | | | |
| No change | 97 | 96 | 82 | 90 | 92 |
| Improve | 1 | 2 | 4 | 4 | 3 |
| Worsen | 1 | 1 | 6 | 2 | 2 |
| Fluctuate | 1 | 2 | 8 | 4 | 3 |
| Recalls current year (b1b) | | | | | |
| No change | 91 | 93 | 87 | 89 | 90 |
| Improve | 6 | 2 | 7 | 4 | 4 |
| Worsen | 2 | 2 | 4 | 4 | 3 |
| Fluctuate | 1 | 2 | 1 | 3 | 2 |
| Recalls current month (b1c) | | | | | |
| No change | 92 | 89 | 90 | 87 | 89 |
| Improve | 5 | 7 | 3 | 4 | 5 |
| Worsen | 2 | 2 | 1 | 6 | 4 |
| Fluctuate | 1 | 2 | 6 | 3 | 3 |
| Recalls current day of week (b1d) | | | | | |
| No change | 74 | 83 | 70 | 66 | 72 |
| Improve | 8 | 7 | 9 | 11 | 9 |
| Worsen | 8 | 5 | 16 | 11 | 10 |
| Fluctuate | 9 | 6 | 6 | 12 | 9 |

| Data Element | HHA (*n* = 95) | IRF (*n* = 122) | LTCH (*n* = 71) | SNF (*n* = 192) | Overall (*n* = 480) |
|---|---|---|---|---|---|
| Recalls "sock" (b1e) | | | | | |
| No change | 77 | 74 | 81 | 67 | 73 |
| Improve | 13 | 14 | 7 | 20 | 15 |
| Worsen | 3 | 6 | 3 | 6 | 5 |
| Fluctuate | 7 | 6 | 9 | 7 | 7 |
| Recalls "blue" (b1f) | | | | | |
| No change | 84 | 79 | 73 | 70 | 75 |
| Improve | 7 | 8 | 16 | 12 | 11 |
| Worsen | 3 | 5 | 1 | 8 | 5 |
| Fluctuate | 5 | 8 | 10 | 10 | 9 |
| Recalls "bed" (b1g) | | | | | |
| No change | 70 | 75 | 67 | 56 | 65 |
| Improve | 14 | 13 | 20 | 19 | 17 |
| Worsen | 10 | 3 | 4 | 8 | 7 |
| Fluctuate | 6 | 8 | 9 | 17 | 11 |
| BIMS impairment category (based on responses to b1a–b1g) | | | | | |
| No change | 84 | 83 | 79 | 72 | 78 |
| Improve | 10 | 9 | 12 | 15 | 12 |
| Worsen | 2 | 1 | 1 | 8 | 4 |
| Fluctuate | 4 | 7 | 7 | 5 | 6 |

### *Admission to Discharge*

Table 3.7 summarizes patterns of change for BIMS data elements from admission to discharge. As described in more detail in Volume 3, discharge data were collected on a subset of the National Beta Test admission sample of patients/residents. Availability of discharge data depended on advance notification of discharge and the ability to schedule assessments among the facility staff assessors at each participating site. Patterns are characterized as "no change" (scores stay the same at admission and discharge), "improve" (scores improve from admission to discharge), and "worsen" (scores decline from admission to discharge). For the BIMS, both admission and discharge data were collected on 794 patients/residents: 146 in HHAs, 338 in IRFs, 84 in LTCHs, and 226 in SNFs. Overall, responses to the BIMS were very similar from admission to discharge. Between 75 percent and 94 percent of scores did not change from admission to discharge. BIMS impairment categorization also tended to be fairly similar at admission and discharge overall (no change in 85 percent, improvement in 11 percent, and decline in 5 percent).

**Table 3.7. Admission to Discharge Results for BIMS Data Elements (percent)**

| Data Element | HHA (*n* = 146) | IRF (*n* = 338) | LTCH (*n* = 84) | SNF (*n* = 226) | Overall (*n* = 794) |
|---|---|---|---|---|---|
| Number of words repeated after first attempt (b1a) | | | | | |
| No change | 90 | 94 | 94 | 95 | 94 |
| Improve | 8 | 4 | 4 | 2 | 4 |
| Worsen | 1 | 2 | 2 | 3 | 2 |
| Recalls current year (b1b) | | | | | |
| No change | 95 | 94 | 88 | 92 | 93 |
| Improve | 4 | 4 | 6 | 4 | 4 |
| Worsen | 1 | 2 | 6 | 4 | 3 |
| Recalls current month (b1c) | | | | | |
| No change | 93 | 94 | 85 | 93 | 92 |
| Improve | 6 | 5 | 10 | 4 | 5 |
| Worsen | 1 | 2 | 6 | 4 | 3 |
| Recalls current day of week (b1d) | | | | | |
| No change | 88 | 84 | 76 | 83 | 84 |
| Improve | 7 | 11 | 12 | 9 | 10 |
| Worsen | 4 | 5 | 12 | 8 | 6 |
| Recalls “sock” (b1e) | | | | | |
| No change | 83 | 81 | 75 | 83 | 81 |
| Improve | 13 | 12 | 19 | 12 | 13 |
| Worsen | 4 | 7 | 6 | 5 | 6 |
| Recalls “blue” (b1f) | | | | | |
| No change | 87 | 84 | 72 | 83 | 83 |
| Improve | 10 | 10 | 18 | 10 | 11 |
| Worsen | 3 | 6 | 10 | 6 | 6 |
| Recalls “bed” (b1g) | | | | | |
| No change | 77 | 77 | 69 | 72 | 75 |
| Improve | 16 | 16 | 17 | 20 | 17 |
| Worsen | 7 | 7 | 14 | 8 | 8 |
| BIMS impairment category (based on responses to b1a–b1g) | | | | | |
| No change | 90 | 87 | 75 | 82 | 85 |
| Worsen | 2 | 4 | 9 | 6 | 5 |
| Improve | 8 | 9 | 16 | 12 | 11 |

When significant change for the overall sample did occur, it reflected patients’/residents’ improved performance in recalling the three words from admission to discharge. Specifically, at discharge (compared with admission), patients/residents overall were 1.75 times more likely to

recall "sock" $\left(t_{(785)} = 4.50, p < 0.001\right)$, 1.52 times more likely to recall "blue" $\left(t_{(785)} = 3.14, p < 0.001\right)$, and 1.78 times more likely to recall "bed" $\left(t_{(787)} = 5.64, p < 0.001\right)$. Furthermore, these improvements in recall contributed to overall performance scores resulting in significant improvements in both BIMS composite score (mean scores not shown in table; $t_{(742)} = 6.27, p < 0.001$) and BIMS categorization $\left(t_{(742)} = 7.70, p < 0.001\right)$ from admission to discharge.

### *Assessor Feedback*

Facility/agency staff indicated that the BIMS has widely accepted clinical utility to track cognition over time and across facilities. When asked to rate the data elements on a five-point scale in the assessor survey, facility/agency staff assessors rated the BIMS among the top five data elements across all the data elements that were tested in the National Beta Test in terms of clinical utility. In the survey, facility/agency staff and research nurses also rated the BIMS as one of the data elements with the lowest burden to both assessors and patients/residents.

In focus groups, research nurses commented favorably on the brevity of the BIMS and facility/agency staff stated that the BIMS is already in widespread use, so it might be a good candidate for further standardization. Assessors also discussed the cue and recall data element as a potential weakness of the BIMS. Research nurses noted the importance of clearly articulating the words that patients/residents will need to repeat (i.e., "bed" versus "red," "sock" versus "suck"). Participants further noted that clear articulation is particularly hard for staff who speak English as a second language, and this issue has the potential to increase burden and affect the validity of the results.

Also of note, facility/agency staff and research nurses felt that the cue and recall section (data elements b1e, b1f, and b1g in Figure 3.1), which asks the patient/resident to recall the words "bed," "sock," and "blue" without or with verbal cues, does not accurately reflect recall abilities. The BIMS word recall could be inadvertently assessing patients'/residents' long-term memory because they had memorized the words when asked the BIMS on a regular basis:

> If they're in a nursing home, skilled facility more than a year, that means at least four times within that year you've asked that same question over again.
>
> —Durham, N.C., SNF staff

Some facility/agency staff suggested overcoming this issue by rotating the three cue and recall words (e.g., shirt, orange, and chair rather than sock, blue, and bed). They recommended choosing a list of similar alternative words representing objects likely to be in the room that even patients/residents who cannot articulate the words easily could point to as a way of indicating that their recall is intact. However, an alternative perspective is that these responses could be informative longitudinally by potentially showing improvement or decline, even if it could be somewhat indicative of long-term rather than short-term memory.

## Summary

Results for the BIMS indicate very high overall support for cross-setting standardization. Assessors noted widespread existing use of the BIMS across all four PAC settings, suggesting minimal increased burden for cross-setting standardization. However, feedback from the assessors indicates that the patients/residents who are assessed with the BIMS frequently tend to remember the three recall words (bed, sock, blue), thus making that task more a test of long-term memory rather than short-term recall ability. Results of the Day 3, 5, and 7 repeat assessments indicate that the BIMS could be administered at any of these points; that is, timing had very little effect on the results for most data elements. Although responses of patients/residents did tend to improve over time on two of the three cue and recall data elements, responses were very similar across the repeat assessments for the other data elements and therefore the improvement could be because of improvement from frequent assessment rather than improved cognitive impairment. However, this result was somewhat an artifact of the National Beta Test design. In practice, the BIMS would not likely be administered frequently enough for this problem to manifest. BIMS categorization was fairly stable from admission to discharge. In addition, the associations between BIMS impairment category and patient/resident characteristics aligned well with our expected results, indicating validity of the data elements. These combined results show high feasibility, excellent interrater reliability, reasonably low burden, and high clinical utility for the BIMS.

# 4. Confusion Assessment Method (CAM)

## Data Element Description

The CAM is an instrument that screens for overall cognitive impairment and includes features to distinguish delirium or reversible confusion from other types of cognitive impairment. Specifically, the CAM screens for change in mental status, inattention, disorganized thinking, and altered level of consciousness. Delirium occurs in up to 50 percent of patients/residents in PAC.[36] Signs and symptoms of delirium are associated with poor functional recovery,[37] rehospitalization, and mortality.[38] Assessment and detection of delirium is essential for identifying and treating the cause of delirium.

The CAM is completed by observing patient/resident behavior during the BIMS data elements for signs and symptoms of delirium, fluctuations in behavior, behaviors that occurred during the assessment period that were not observed during the BIMS interview, reviewing medical record documentation to determine cognitive function at the time of initial assessment, and interviewing staff, family members, and others who observe the patient's/resident's behavior. The CAM is under separate copyright protection.[39] The Hospital Elder Life Program has granted permission to use the CAM in association with the PAC instruments. Versions of the CAM are currently in the MDS 3.0 and LCDS. In the National Beta Test, the CAM was included in the repeat assessment evaluation and was administered repeatedly to a subset of the sample on the same patient/resident by the same assessor on Days 3, 5, and 7. The CAM is shown in Figure 4.1.

[36] Kiely et al., 2004; Marcantonio et al., 2003.

[37] Marcantonio et al., 2003.

[38] Marcantonio et al., 2005.

[39] Adapted from Inouye et al., 1990. 

**Figure 4.1. Signs and Symptoms of Delirium (from CAM)**

**INSTRUCTIONS: CODE ONLY AFTER COMPLETING THE BRIEF INTERVIEW FOR MENTAL STATUS (B1).**

**B2a. Acute Onset Mental Status Change:** Is there evidence of an acute change in mental status from the patient's/resident's baseline?

- ☐ 0 = No
- ☐ 1 = Yes

**B2b. Inattention:** Did the patient/resident have difficulty focusing attention, for example, being easily distracted or having difficulty keeping track of what was being said?

- ☐ 0 = Behavior not present
- ☐ 1 = Behavior continuously present, does not fluctuate
- ☐ 2 = Behavior present, fluctuates (comes and goes, changes in severity)

**B2c. Disorganized Thinking:** Was the patient's/resident's thinking disorganized or incoherent (rambling or irrelevant conversation, unclear or illogical flow of ideas, or unpredictable switching from subject to subject)?

- ☐ 0 = Behavior not present
- ☐ 1 = Behavior continuously present, does not fluctuate
- ☐ 2 = Behavior present, fluctuates (comes and goes, changes in severity)

**B2d. Altered Level of Consciousness:** Did the patient/resident have altered level of consciousness, as indicated by ANY of the following criteria?

- **vigilant** – startled easily to any sound or touch
- **lethargic** – repeatedly dozed off when being asked questions, but responded to voice or touch
- **stuporous** – very difficult to arouse and keep aroused for the interview
- **comatose** – could not be aroused

- ☐ 0 = Behavior not present
- ☐ 1 = Behavior continuously present, does not fluctuate
- ☐ 2 = Behavior present, fluctuates (comes and goes, changes in severity)

1. Adapted from Inouye et al., 1990. The Confusion Assessment Method (CAM) is the property of Hospital Elder Life Program. 

## Testing Objectives

Basic descriptive statistics (i.e., frequencies) are presented for CAM admission data. To examine known groups' validity, we also examined "evidence of change in mental status from baseline" by patient/resident characteristics and clinical groups of interest. For admission data, feasibility (rates of missingness and time to complete) and interrater reliability (kappa and percent agreement) were examined. The CAM was administered repeatedly on the same patient/resident by the same assessor to test the effects of conducting the assessment on Day 3, 5, or 7. As such, an additional objective was to understand whether there were significant and meaningful differences in rates or scores depending on the day a patient/resident was assessed. Lastly, frequencies at admission and discharge were compared to inform stability or possible change over time.

## Results

### *Feasibility*

#### Frequencies/Missing

Table 4.1 shows the percent of responses at admission for each CAM data element overall and by setting. The CAM was administered to 2,973 patients/residents: 630 in HHAs, 771 in IRFs, 471 in LTCHs, and 1,101 in SNFs. Just over 95 percent of the sample were administered the CAM at admission. Among these, overall missing data at the data element level ranged from 0.1 percent to 0.4 percent, with minimal setting differences. Overall, 5 percent of patients/residents had "evidence of change in mental status from baseline," 12 percent had "difficultly focusing," 6 percent had "disorganized thinking," and 4 percent had "altered consciousness." Setting-specific results indicated that the frequency of "change in mental status from baseline" ranged from 4 percent in SNFs to 6 percent in IRFs. Rates of patients/residents with "difficulty focusing" ranged from 10 percent in SNFs to 15 percent in IRFs. Rates of those with "disorganized thinking" ranged from 5 percent in HHAs to 7 percent in LTCHs, and those with "altered consciousness" ranged from 2 percent in HHAs to 6 percent in LTCHs.

**Table 4.1. Overall and Setting-Specific Response Frequencies for CAM Data Elements at Admission (percent)**

| Data Element | HHA (*n* = 630) | IRF (*n* = 771) | LTCH (*n* = 471) | SNF (*n* = 1,101) | Overall (*n* = 2,973) |
|---|---|---|---|---|---|
| Evidence of change in mental status from baseline (b2a) | | | | | |
| Yes | 5 | 6 | 5 | 4 | 5 |
| Did patient have difficulty focusing attention (b2b) | | | | | |
| Behavior not present | 89 | 85 | 89 | 90 | 88 |
| Behavior continuously present | 2 | 3 | 3 | 3 | 3 |
| Behavior present, fluctuates | 9 | 11 | 8 | 8 | 9 |
| Was patient thinking disorganized (b2c) | | | | | |
| Behavior not present | 95 | 94 | 93 | 94 | 94 |
| Behavior continuously present | 1 | 2 | 2 | 1 | 1 |
| Behavior present, fluctuates | 4 | 5 | 4 | 6 | 5 |
| Did patient have altered consciousness (b2d) | | | | | |
| Behavior not present | 98 | 95 | 94 | 96 | 96 |
| Behavior continuously present | 1 | 1 | 2 | 1 | 1 |
| Behavior present, fluctuates | 2 | 3 | 3 | 3 | 3 |

### Known Groups Validity

Comparing the performance of patients/residents on the CAM with other patient/resident characteristics adds information about the validity of data elements. If known or logical associations between patients/resident characteristics and data elements are observed in data from the National Beta test, this information contributes to the evidence that the data elements are valid, or assessing the construct that they are intended to capture.

Table 4.2 shows the breakdown for "evidence of change in mental status from baseline" for the overall admission sample, stratified by patient/resident characteristics and clinical groups as described in Chapter 1: gender (male or female, as documented by National Beta Test assessor), age (as categorized into the following ranges: 18–44, 45–64, 65–74, 75–89, 90 and over), length of stay (in days), disposition at discharge (e.g., to another PAC setting, home, to hospital), sepsis, heart failure, stroke, and two ADLs—toileting (not available for HHA patients) and ability to transfer from lying to sitting. As a reminder, these clinical conditions were chosen based on their common occurrence across settings, their frequent relationship with many of the data elements tested in the National Beta Test, and their availability in all four standardized assessments (OASIS, IRF-PAI, LCDS, MDS). Setting-specific results are presented in the appendix (Tables A.13–A.16).

**Table 4.2. Frequencies for the CAM Change in Mental Status Data Element by Patient/Resident Characteristics and Clinical Groups (percent)**

| Patient/Resident Characteristics and Clinical Groups | Change in Mental Status (Yes) |
|---|---|
| Gender (*n* = 2,862[a]) | |
| Male (*n* = 1,179) | 6.1 |
| Female (*n* = 1,683) | 3.9 |
| Age (*n* = 2,852) | |
| 18–44 (*n* = 37) | 0.0 |
| 45–64 (*n* = 304) | 5.3 |
| 65–74 (*n* = 900) | 4.9 |
| 75–89 (*n* = 1,306) | 4.9 |
| 90 or older (*n* = 305) | 3.9 |
| Length of stay (*n* = 2,530; mean, SD) | Yes: 20.8 (11.3)<br>No: 21.6 (12.8) |
| Disposition at discharge (*n* = 2,817[a]) | |
| Home (*n* = 1,315) | 4.7 |
| Hospital (*n* = 191) | 4.2 |
| Hospice (*n* = 40) | 12.5 |
| SNF (*n* = 274) | 6.2 |
| IRF (*n* = 51) | 5.9 |
| HHA (*n* = 615) | 2.6 |
| LTCH (*n* = 12) | 8.3 |
| Other (*n* = 319) | 6.9 |
| Clinical conditions (*n* = 2,199) | |
| Sepsis | |
| Yes (*n* = 148) | 6.1 |
| No (*n* = 2,051) | 5.0 |
| Heart failure | |
| Yes (*n* = 373) | 6.2 |
| No (*n* = 1,826) | 4.8 |
| Stroke | |
| Yes (*n* = 192[a]) | 10.9 |
| No (*n* = 2,007) | 4.5 |

| Patient/Resident Characteristics and Clinical Groups | Change in Mental Status (Yes) |
|---|---|
| Hygiene—Toileting (*n* = 1,489)[b] | |
| Independent (*n* = 72) | 2.8 |
| Setup or clean-up assistance (*n* = 76) | 9.2 |
| Supervision or touching assistance (*n* = 315) | 3.5 |
| Partial/moderate assistance (*n* = 357) | 4.8 |
| Substantial/maximal assistance (*n* = 330) | 5.8 |
| Dependent (*n* = 339) | 7.7 |
| Mobility—Transfer from lying to sitting (*n* = 1,838) | |
| Independent (*n* = 191) | 3.7 |
| Setup or clean-up assistance (*n* = 110) | 8.2 |
| Supervision or touching assistance (*n* = 519) | 4.1 |
| Partial/moderate assistance (*n* = 608) | 5.4 |
| Substantial/maximal assistance (*n* = 282) | 5.3 |
| Dependent (*n* = 128) | 9.4 |

[a] Significant ($p < 0.05$) associations with change in mental status as indicated by chi-square tests of independence (ANOVA for length of stay).
[b] Toileting data not available for HHA patients.

Based on the research literature, we generated several hypotheses or expectations for associations between the data elements and the patient/resident characteristics. We expected a change in mental status as assessed by the CAM to be related to age and stroke, such that those patients/residents displaying a change in mental status would tend to be older[40] and more likely to have suffered a stroke.[41]

Across the full sample, we observed significant associations between change in mental status from baseline and gender, disposition at discharge, and stroke. There were no other significant associations.

*Gender and Age*

- Gender was associated with change in mental status overall $\left(\chi^2_{(1)} = 7.66, p < 0.01\right)$, with greater change in mental status observed in males (6 percent) than in females (4 percent). At the setting level, however, gender was not associated with change in mental status in any of the four settings. *This association conforms with related findings suggesting men could be at greater risk of delirium in some clinical*

[40] Inouye, 2006.

[41] Shi et al., 2012.

*situations,*[42] *or it could be because of a third variable that differs between men and women, such as age, cognitive function, or use of certain medications.*[43]

- There was no overall association between age and change in mental status. For most age groups, between 4 percent and 5 percent demonstrated a change in mental status from baseline. Similarly, age was not associated with change in mental status in any of the four settings. *This lack of association is contrary to our expectation, as older age has been shown to be a risk factor for delirium.*[44] *It is possible that the relatively low rates of patients/residents with a change in mental status, combined with the specific age categories, contribute to this lack of observed effect.*

*Length of Stay and Disposition at Discharge*

- Length of stay was similar for those with and without a change in mental status: 20.8 days (SD = 11.3) and 21.6 days (SD = 12.8), respectively. At the setting level, however, length of stay was significantly associated with change in mental status only in IRFs, where those with a change in mental status had a significantly longer length of stay than those without. *This finding is consistent with some prior studies that have found associations between length of stay and delirium.*[45]
- Disposition at discharge was associated with change in mental status $\left(\chi^2_{(7)} = 16.71, p < 0.05\right)$. A greater percentage of patients/residents being discharged to hospice (13 percent) and LTCHs (8 percent) demonstrated evidence of a change in mental status, while, for instance, only 3 percent of patients/residents discharged to HHAs demonstrated evidence of a change in mental status. At the setting level, however, disposition at discharge was not associated with a change in mental status in any of the four settings. *These findings are difficult to interpret but somewhat consistent with findings that patients/residents with delirium are sicker overall and have poorer prognosis than those without delirium.*[46]

*Clinical Conditions*

- Frequencies show that, overall, 5 percent of those with and without sepsis had a change in mental status from baseline, indicating no significant associations between change in mental status and sepsis. Likewise, no association was observed by setting. *This is consistent with our expectation of no relationship between change in mental status and this clinical condition*.
- There was also no association between heart failure and change in mental status. Six percent of those with heart failure and 5 percent of those without had a change in mental status from baseline. Likewise, no association was observed by setting. *This*

---

[42] Oh et al., 2016; Lee et al., 2011; Edlund et al., 2001.

[43] Lundstrom et al., 2007.

[44] Inouye, 2006.

[45] Marcantonio et al., 1994.

[46] Edlund et al, 2001; Lundstrom et al., 2007.

*is consistent with our expectation of no relationship between change in mental status and this clinical condition.*

- There was, however, a significant association between stroke and change in mental status $\left(\chi^2_{(1)} = 15.8, p < 0.01\right)$: 11 percent of stroke and 5 percent of nonstroke patients/residents showed evidence of a change in mental status from baseline. This association was also evident among patients/residents in IRFs and SNFs but not in LTCHs or HHAs. *This finding is consistent with our expectation and with prior literature, and contributes to evidence of the validity of the CAM for use in this population.*

*ADLs: Toileting and Ability to Transfer from Lying to Sitting*

- Change in mental status was not associated with either of the ADLs, overall or at the setting level. *This is consistent with our expectation of no relationship between change in mental status and ADLs.*

### Time to Complete

Table 4.3 shows average time to complete the CAM overall and by setting at admission. Overall, the mean time to complete the CAM was 1.4 minutes (SD = 0.7 minutes). Setting-specific time to complete ranged from 1.3 minutes (SD = 0.6 minutes) in IRFs to 1.5 minutes (SD = 0.7 minutes) in LTCHs and HHAs. Time to complete was associated with setting type $\left(F_{(3,1532)} = 11.21, p < 0.0001\right)$: It took less time to complete in IRFs than in LTCHs $\left(t_{(1532)} = 5.24, p < 0.001\right)$, HHAs $\left(t_{(1532)} = 4.28, p < 0.001\right)$, or SNFs $\left(t_{(1532)} = 3.62, p < 0.001\right)$. There were no significant differences among HHAs, LTCHs, and SNFs.

**Table 4.3. Time to Complete the CAM Data Elements (minutes)**

| Characteristic | HHA (*n* = 375) | IRF (*n* = 472) | LTCH (*n* = 284) | SNF (*n* = 405) | Overall (*n* = 1,536) |
|---|---|---|---|---|---|
| Mean (SD) | 1.5 (0.7) | 1.3 (0.6)[a] | 1.5 (0.7) | 1.4 (0.7) | 1.4 (0.7) |

[a] Significantly lower time to complete in IRFs than in any other setting.

Time to complete was also evaluated according to urbanicity (urban versus nonurban), geographic region (Northeast, South, Midwest, West), facility ownership (for-profit versus nonprofit), and facility size (above or below setting-type median) to evaluate the generalizability of these performance results (see Tables A.17–A.20 in the appendix). Assessments were completed more quickly in urban than in nonurban settings (effect size = 0.43). Also, the Midwest time to complete of 1.5 minutes was significantly longer than the Northeast region (1.4 minutes), but the effect size was very small (0.11). It also took significantly more time (1.5 minutes) for smaller facilities to complete this section than larger facilities (1.3 minutes), but the effect size was small (0.28). No other significant differences were found for time to complete the CAM in these sensitivity analyses.

### *Interrater Reliability*

Table 4.4 shows kappa interrater reliability coefficients for the CAM overall and by setting. As described in more detail in Volume 3, paired assessment data for interrater reliability evaluation were collected on a subset of the National Beta Test admission sample of patients/residents according to setting-level target totals. For example, each participating LTCH was asked to conduct 20 paired assessments to contribute to interrater reliability. Inclusion in interrater reliability data collection depended on paired facility staff and research nurse assessors' ability to schedule assessments. Kappa coefficients were computed on data obtained from 914 patients/residents: 189 in HHAs, 245 in IRFs, 223 in LTCHs, and 257 in SNFs. Only the "difficulty focusing" data element yielded a stable overall kappa coefficient, which fell into the range of good agreement at 0.66. Kappa coefficients for this data element at the setting level were also good, except in IRFs, where kappa was moderate (0.55). Kappas were also good for the "change in mental status from baseline" data element (0.60) in IRFs and the "disorganized thinking" data element (0.68) in SNFs. As a reminder, because of the impact of prevalence rates on the stability and interpretability of kappa estimates, kappa is not reported for data elements with prevalence rates out of range for stable kappa estimates.

Interrater reliability (kappa) was also evaluated according to urbanicity (urban versus nonurban), geographic region (Northeast, South, Midwest, West), facility ownership (for-profit versus nonprofit), and facility size (above or below setting-type median) to evaluate the generalizability of these performance results (see Tables A.21–A.24 in the appendix). No noteworthy differences were found for interrater reliability of the CAM in these sensitivity analyses.

**Table 4.4. Interrater Reliability Kappa or Weighted Kappa for CAM Data Elements**

| Data Element | HHA (*n* = 189) | IRF (*n* = 245) | LTCH (*n* = 223) | SNF (*n* = 257) | Overall (*n* = 914) |
|---|---|---|---|---|---|
| Evidence of change in mental status from baseline (b2a) | — | 0.60 | — | — | — |
| Did patient have difficulty focusing attention (b2b) | 0.66 | 0.55 | 0.75 | 0.70 | 0.66 |
| Was patient thinking disorganized (b2c) | — | — | — | 0.68 | — |
| Did patient have altered consciousness (b2d) | — | — | — | — | — |

NOTES: Interrater reliability not shown for data elements with proportions out of range for stable kappa estimate (per study power calculations). Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

Table 4.5 shows percent agreement reliability for the CAM data elements overall and by setting. Overall, percent agreement was high for all data elements, ranging from 91 percent to 96 percent. At the setting level, percent agreement was high across settings for all data elements, ranging from 94 percent to 98 percent in HHAs, 89 percent to 97 percent in IRFs, 93 percent to 98 percent in LTCHs, and 93 percent to 97 percent in SNFs.

**Table 4.5. Interrater Reliability—Percent Agreement for CAM Data Elements**

| Data Element | HHA (*n* = 189) | IRF (*n* = 245) | LTCH (*n* = 223) | SNF (*n* = 257) | Overall (*n* = 914) |
|---|---|---|---|---|---|
| Evidence of change in mental status from baseline (b2a) | 97 | 93 | 98 | 97 | 96 |
| Did patient have difficulty focusing attention (b2b) | 91 | 89 | 93 | 93 | 91 |
| Was patient thinking disorganized (b2c) | 94 | 93 | 96 | 94 | 94 |
| Did patient have altered consciousness (b2d) | 98 | 97 | 95 | 96 | 96 |

### *Day 3, 5, 7 Repeat Assessment Evaluation*

Table 4.6 summarizes patterns of change for the CAM across the repeat assessment days. Patterns are characterized as "no change" (scores stay the same across assessment days), "improve" (steady improvement across assessment days), "worsen" (steady decline across assessment days), and "fluctuate" (scores go up and down across assessment days). As described in more detail in Volume 3, repeat assessment data were collected on a subset of the National Beta Test admission sample of patients/residents according to setting-level target totals. For example, each participating HHA was asked to contribute five patients for repeat assessment. Inclusion in repeat assessment data collection depended on assessor ability to complete the initial assessment on Day 3, availability of the assessor to return for repeat assessments on Days 5 and 7, and willingness of patient/resident to complete multiple assessments. For the CAM, 475 patients/residents participated in repeat assessments at Days 3, 5, and 7 from admission: 94 in HHAs, 121 in IRFs, 71 in LTCHs, and 189 in SNFs. Scores tended to remain relatively consistent across the repeat assessment days. Overall, 91 percent of patients/residents had consistent assessments regarding change in mental status from baseline for the three assessment days, 87 percent had consistent assessments regarding difficulty focusing, 90 percent had consistent assessments regarding disorganized thinking, and 95 percent had consistent assessments regarding altered consciousness. Consistent with these trends, there were no statistically significant overall differences across assessment days on any of the CAM data elements.

**Table 4.6. Day 3, 5, and 7 Repeat Assessment Results for CAM Data Elements (percent)**

| Data Element | HHA (*n* = 94) | IRF (*n* = 121) | LTCH (*n* = 71) | SNF (*n* = 189) | Overall (*n* = 475) |
|---|---|---|---|---|---|
| Evidence of change in mental status from baseline (b2a) | | | | | |
| No change | 94 | 90 | 87 | 93 | 91 |
| No to Yes | 2 | 3 | 6 | 2 | 3 |
| Yes to No | 3 | 5 | 4 | 3 | 4 |
| Fluctuate | 1 | 3 | 3 | 2 | 2 |
| Did patient have difficulty focusing attention (b2b)[a] | | | | | |
| No change | 85 | 87 | 86 | 88 | 87 |
| Worsen | 6 | 6 | 3 | 2 | 4 |
| Improve | 5 | 6 | 6 | 7 | 6 |
| Fluctuate | 3 | 2 | 6 | 3 | 3 |
| Was patient thinking disorganized (b2c)[a] | | | | | |
| No change | 93 | 91 | 90 | 89 | 90 |
| Worsen | 0 | 3 | 4 | 1 | 2 |
| Improve | 4 | 4 | 4 | 5 | 4 |
| Fluctuate | 3 | 2 | 1 | 5 | 3 |
| Did patient have altered consciousness (b2d)[a] | | | | | |
| No change | 98 | 96 | 88 | 95 | 95 |
| Worsen | 2 | 2 | 4 | 1 | 2 |
| Improve | 0 | 2 | 6 | 3 | 3 |
| Fluctuate | 0 | 1 | 1 | 1 | 1 |

[a] To evaluate change, data elements were first recoded to combine the two categories representing "behavior present"—"behavior continuously present, does not fluctuate" and "behavior present, fluctuates (comes and goes, changes in severity)"—into a single category, resulting in dichotomized data elements (behavior present/behavior nor present) with which to evaluate change.

### *Admission to Discharge*

Table 4.7 summarizes patterns of change on the CAM from admission to discharge. With the exception of data element b2a, patterns are characterized as "no change" (scores stay the same at admission and discharge), "improve" (scores indicate improvement from admission to discharge), and "worsen" (scores indicate decline from admission to discharge). As described in more detail in Volume 3, discharge data were collected on a subset of the National Beta Test admission sample of patients/residents. Availability of discharge data depended on advance notification of discharge and the ability to schedule assessments among the facility staff assessors at each participating site. For the CAM, both admission and discharge data were collected on 775 patients/residents: 140 in HHAs, 332 in IRFs, 82 in LTHCs, and 221 in SNFs. Overall, responses to the CAM data elements were similar from admission to discharge: 88

percent to 96 percent of scores remained the same from admission to discharge. When change was present, it tended to reflect improvement from admission to discharge. For instance, regarding "difficulty focusing," 8 percent of patients/residents overall showed improvement at discharge. There were no statistically significant overall differences in responses between admission and discharge.

**Table 4.7. Admission to Discharge Results for CAM Data Elements (percent)**

| Data Element | HHA (*n* = 140) | IRF (*n* = 332) | LTCH (*n* = 82) | SNF (*n* = 221) | Overall (*n* = 775) |
|---|---|---|---|---|---|
| Evidence of change in mental status from baseline (b2a) | | | | | |
| No change | 96 | 95 | 88 | 96 | 95 |
| No to Yes | 1 | 1 | 5 | 1 | 2 |
| Yes to No | 3 | 4 | 7 | 2 | 4 |
| Did patient have difficulty focusing attention (b2b)[a] | | | | | |
| No change | 87 | 85 | 87 | 95 | 88 |
| Worsen | 3 | 5 | 9 | 3 | 4 |
| Improve | 10 | 11 | 5 | 3 | 8 |
| Was patient thinking disorganized (b2c)[a] | | | | | |
| No change | 95 | 94 | 93 | 96 | 95 |
| Worsen | 3 | 2 | 5 | 2 | 2 |
| Improve | 2 | 4 | 2 | 2 | 3 |
| Did patient have altered consciousness (b2d)[a] | | | | | |
| No change | 97 | 94 | 98 | 98 | 96 |
| Worsen | 1 | 2 | 1 | 1 | 1 |
| Improve | 1 | 4 | 1 | 1 | 2 |

[a] To evaluate change, data elements were first recoded to combine the two categories representing "behavior present"—"behavior continuously present, does not fluctuate" and "behavior present, fluctuates (comes and goes, changes in severity)"—into a single category, resulting in dichotomized data elements (behavior present/behavior nor present) with which to evaluate change.

### *Assessor Feedback*

According to the assessor survey, facility/agency staff and research nurses found the CAM to be moderately clinically useful, although there was some variation in opinion about the overall usefulness of the CAM for cross-setting use. Specifically, facility assessors in LTCHs rated the CAM's clinical utility higher (average score = 4.2 out of 5) than did facility assessors in other settings (SNFs = 3.8, HHAs = 3.4, IRFs = 3.3). Other data collection (e.g., focus groups) and interaction with the field test sites did not offer an explanation for this difference. Given that

delirium is common in patients who have recently had surgery or experienced trauma,[47] and that these patients are likely to be overrepresented in LTCHs compared with other PAC settings, we believe that this difference in rating may be because of a higher prevalence of delirium in LTCHs. Facility/agency staff and research nurses considered the CAM to be a relatively low-burden data element. The assessors did not offer any additional feedback in the focus groups regarding the CAM.

## Summary

Results for the CAM indicate reasonable overall support for cross-setting standardization. Results from the National Beta Test support feasibility and interrater reliability. Burden was considered to be minimal, and the CAM showed moderate clinical utility. Results of the Day 3, 5, and 7 repeat assessment indicate that the CAM was stable regardless of the day the assessment was administered; that is, responses were similar across the three days the assessment was administered. The CAM was also stable from admission to discharge. In addition, the associations between change in mental status and patient/resident characteristics somewhat aligned with our expected results, indicating validity of the data elements.

[47] Mehta et al., 2015; Pandharipande et al., 2008.

# 5. Expression and Understanding

## Data Element Description

The Expression and Understanding data elements assess whether a patient/resident is able to express or communicate requests, needs, and opinions; conduct social conversation; and comprehend direct person-to-person communication. Issues regarding making oneself understood (ability to communicate requests and needs) and inability to understand communication can be very frustrating and can contribute to social isolation,[48] poor mood, and behavior disorders.[49] Assessment of a patient's/resident's ability to communicate and understand others can help identify underlying causes and the best methods to facilitate communication and understanding for the patient/resident.

Currently, the data elements Expression of Ideas and Wants and Understanding Verbal Content are included in the IRF-PAI and LCDS. The MDS includes similar data elements (Makes Self Understood, Ability to Understand Others, and Speech Clarity). In the National Beta Test, both the two–data element (without Speech Clarity) LCDS/IRF-PAI version and the three–data element (with Speech Clarity) MDS version were tested. In addition, both versions were included in the repeat assessment evaluation and therefore admission data were collected repeatedly on the same patient/resident by the same assessor on Days 3, 5, and 7. Figure 5.1 shows the three–data element (MDS) version, and Figure 5.2 shows the two–data element (LCDS/IRF-PAI) version.

## Testing Objectives

Two versions of the Expression and Understanding data elements were administered during the National Beta Test: a three–data element (MDS) version and a two–data element (LCDS/IRF-PAI) version. The three–data element version, which includes Speech Clarity in addition to Expression and Understanding, was administered in Market Group A, and the two–data element version, which does not include a data element for Speech Clarity, was administered in Market Group B. For more details on Market Group A and B samples and characteristics, see Volume 2. All analyses described were performed on both versions and were conducted overall and by PAC setting.

[48] Resnick, Fries, and Verbrugge, 1997.

[49] Beck, Rossby, and Baldwin, 1991.

**Figure 5.1. Three–Data Element Expression of Ideas and Wants and Understanding Verbal Content (MDS Version)**

**A3.** Select best description of speech pattern

- ☐ 0 = Clear speech – distinct intelligible words
- ☐ 1 = Unclear speech – slurred or mumbled words
- ☐ 2 = No speech – absence of spoken words
- ☐ 9 = **Unknown or unable to assess**

**A4.** Ability to express ideas and wants, consider both verbal and non-verbal expression

- ☐ 0 = Understood
- ☐ 1 = Usually understood – difficulty communicating some words or finishing thoughts but is able if prompted or given time
- ☐ 2 = Sometimes understood – ability is limited to making concrete requests
- ☐ 3 = Rarely/never understood
- ☐ 9 = **Unknown or unable to assess**

**A5.** Understanding verbal content, however able (with hearing aid or device if used)

- ☐ 0 = Understands – clear comprehension
- ☐ 1 = Usually understood – misses some part/intent of message but comprehends most conversation
- ☐ 2 = Sometimes understands – responds adequately to simple, direct communication only
- ☐ 3 = Rarely/never understands
- ☐ 9 = **Unknown or unable to assess**

**Figure 5.2. Two–Data Element Expression of Ideas and Wants and Understanding Verbal Content (IRF-PAI and LCDS Version)**

**A6.** Expression of Ideas and Wants (consider both verbal and non-verbal expression and excluding language barriers)

- ☐ 4 = Expresses complex messages without difficulty and with speech that is clear and easy to understand
- ☐ 3 = Exhibits some difficulty with expressing needs and ideas (e.g., some words or finishing thoughts) or speech is not clear
- ☐ 2 = Frequently exhibits difficulty with expressing needs and ideas
- ☐ 1 = Rarely/Never expresses self or speech is very difficult to understand
- ☐ 9 = **Unknown or unable to assess**

**A7.** Understanding Verbal Content (with hearing aid or device, if used and excluding language barriers)

- ☐ 4 = Understands: Clear comprehension without cues or repetitions
- ☐ 3 = Usually Understands: Understands most conversations, but misses some part/intent of message. Requires cues at times to understand
- ☐ 2 = Sometimes Understands: Understands only basic conversation or simple, direct phrases. Frequently requires cues to understand
- ☐ 1 = Rarely/Never Understands
- ☐ 9 = **Unknown or unable to assess**

Basic descriptive statistics (e.g., frequencies) are presented for admission data to characterize the rates of speech clarity, expression, and understanding for patients/residents in each setting and for the overall sample for both data element versions. We also examined the ability to express oneself (i.e., "expression of ideas and wants" and "makes self understood") and understand others (i.e., "understanding verbal content" and "ability to understand others") by patient/resident characteristics and clinical groups of interest. For these analyses, data from the two versions were combined. For admission data, feasibility (rates of missingness and time to complete) and interrater reliability (kappa and percent agreement) were examined separately for both versions. Both versions of the Expression and Understanding data elements were administered repeatedly on the same patient/resident by the same assessor to test the effects of conducting the assessment on Day 3, 5, or 7 after admission. As such, an additional objective was to understand whether there were significant and meaningful differences in rates or scores depending on the day a patient/resident was assessed. Lastly, frequencies at admission and discharge were compared to inform stability or possible change over time for both data element versions.

## Results

### *Feasibility*

#### Frequencies/Missing

Tables 5.1 and 5.2 show the percentage of responses for each Expression and Understanding data element overall and by setting. Table 5.1 shows Market Group A, which was administered the three–data element version, and Table 5.2 shows Market Group B, which was administered the two–data element version. The three–data element version was administered to 1,534 patients/residents: 229 in HHAs, 489 in IRFs, 265 in LTCHs, and 551 in SNFs. The two–data element version was administered to 1,529 patients/residents: 412 in HHAs, 295 in IRFs, 233 in LTCHs, and 589 in SNFs. Overall, more than 98 percent of the sample was administered one of the two versions; the section was not completed for only 2 percent of the sample. Among these incompletes, overall missing data at the data element–level ranged from 0 percent to 0.1 percent across the two versions, with minimal setting differences for either version. Results for Speech Clarity are shown only for Market Group A, as an analogous data element was not administered in Market Group B. For the three– and two–data element versions, 90 percent of patients/residents were "understood" or "expressed without difficulty," and 88 percent and 89 percent of patients/residents "understood verbal content" or "understood without cues or repetitions," respectively. For patients/residents who received the three–data element version, 95 percent exhibited clear speech.

**Table 5.1. Overall and Setting-Specific Response Frequencies for Expression and Understanding Data Elements: Market Group A (percent)**

| Data Element | HHA (*n* = 229) | IRF (*n* = 489) | LTCH (*n* = 265) | SNF (*n* = 551) | Overall (*n* = 1,534) |
|---|---|---|---|---|---|
| Speech clarity (a3) | | | | | |
| Clear speech | 96 | 94 | 93 | 97 | 95 |
| Unclear speech | 4 | 6 | 4 | 3 | 5 |
| No speech | 0 | 0 | 3 | 0 | 1 |
| Ability to express ideas and wants (a4) | | | | | |
| Understood | 74 | 94 | 91 | 92 | 90 |
| Usually understood | 24 | 5 | 8 | 7 | 9 |
| Sometimes understood | 1 | 1 | 1 | 1 | 1 |
| Rarely/never understood | 0 | 0 | 0 | 0 | 0 |
| Understanding verbal content (a5) | | | | | |
| Understands | 69 | 93 | 91 | 90 | 88 |
| Usually understands | 31 | 5 | 8 | 9 | 11 |
| Sometimes understands | 1 | 2 | 2 | 1 | 1 |
| Rarely/never understands | 0 | 0 | 0 | 0 | 0 |

**Table 5.2. Overall and Setting-Specific Response Frequencies for Expression and Understanding Data Elements: Market Group B (percent)**

| Data Element | HHA (*n* = 412) | IRF (*n* = 295) | LTCH (*n* = 233) | SNF (*n* = 589) | Overall (*n* = 1,529) |
|---|---|---|---|---|---|
| Expresses ideas and wants (a6) | | | | | |
| Expresses without difficulty | 90 | 87 | 91 | 92 | 90 |
| Exhibits some difficulty | 10 | 11 | 7 | 7 | 9 |
| Frequently exhibits difficulty | 1 | 1 | 2 | 1 | 1 |
| Rarely/never expresses | 0 | 0 | 0 | 0 | 0 |
| Understands verbal content (a7) | | | | | |
| Understands without cues or repetitions | 86 | 88 | 91 | 90 | 89 |
| Usually understands | 13 | 11 | 9 | 9 | 10 |
| Sometimes understands | 1 | 1 | 1 | 1 | 1 |
| Rarely/never understands | 0 | 0 | 0 | 0 | 0 |

### Known Groups Validity

Comparing the performance of patients/residents on the Expression and Understanding data elements with other patient/resident characteristics adds information about validity. If known or logical associations between patients/resident characteristics and data elements are observed in data from the National Beta Test, this information contributes to the evidence that the data elements are valid, or assessing the construct that they are intended to capture.

Table 5.3 shows the frequencies for data elements specific to Expression and Understanding (Speech Clarity excluded) for the overall admission sample, with data from the two market groups combined and stratified by patient/resident characteristics and clinical groups, as described in Chapter 1: gender (male or female, as documented by National Beta Test assessor), age (as categorized into the following ranges: 18–44, 45–64, 65–74, 75–89, 90 and over), length of stay (in days), disposition at discharge (e.g., to another PAC setting, home, to hospital), sepsis, heart failure, stroke, and two ADLs—toileting (not available for HHA patients) and ability to transfer from lying to sitting. As a reminder, these clinical conditions were chosen based on their common occurrence across settings, their frequent relationship with many of the data elements tested in the National Beta Test, and their availability in all four standardized assessments (OASIS, IRF-PAI, LCDS, MDS). Setting-specific results are presented in the appendix (Tables A.25–A.28). Responses were dichotomized to reflect ability to express oneself or not and ability to understand others or not; comparable data elements from both versions (three– and two–data element) were combined. Because of slight differences in available data for the Expression and Understanding data elements, we report sample sizes for each (*nE* = expression; *nU* = understanding).

**Table 5.3. Frequencies for Expression and Understanding Data Elements by Patient/Resident Characteristics and Clinical Groups (percent)**

| Patient/Resident Characteristics and Clinical Groups | Expresses without Difficulty (Yes) | Understands (Yes) |
|---|---|---|
| Gender (*nE* = 2,951; *nU* = 2,950) | | |
| Male (*nE* = 1,220; *nU* = 1,222) | 89.8 | 88.1 |
| Female (*nE* = 1,731; *nU* = 1,728) | 90.2 | 88.4 |
| Age (*nE* = 2,941[a]; *nU* = 2,940[a]) | | |
| 18–44 (*nE* = 42; *nU* = 42) | 88.1 | 88.1 |
| 45–64 (*nE* = 309; *nU* = 310) | 91.6 | 92.6 |
| 65–74 (*nE* = 917; *nU* = 918) | 93.7 | 92.2 |
| 75–89 (*nE* = 1357; *nU* = 1,354) | 88.1 | 86.6 |
| 90 or older (*nE* = 316; *nU* = 316) | 86.7 | 80.1 |
| Length of stay (*nE* = 2,598[a]; *nU* = 2,597[a]; mean, SD) | Yes: 21.2 (12.4)<br>No: 25.2 (15.0) | Yes: 21.0 (12.3)<br>No: 26.2 (15.6) |

| Patient/Resident Characteristics and Clinical Groups | Expresses without Difficulty (Yes) | Understands (Yes) |
|---|---|---|
| Disposition at discharge (*nE* = 2,896[a]; *nU* = 2,895[a]) | | |
| Home (*nE* = 1,346; *nU* = 1,344) | 90.3 | 87.4 |
| Hospital (*nE* = 202; *nU* = 202) | 91.1 | 89.6 |
| Hospice (*nE* = 41; *nU* = 40) | 85.4 | 80.0 |
| HHA (*nE* = 626; *nU* = 626) | 94.9 | 92.8 |
| IRF (*nE* = 51; *nU* = 51) | 92.2 | 92.2 |
| LTCH (*nE* = 13; *nU* = 13) | 76.9 | 100 |
| SNF (*nE* = 288; *nU* = 288) | 86.5 | 88.9 |
| Other (*nE* = 329; *nU* = 331) | 82.4 | 81.9 |
| Clinical conditions (*nE* = 2,272; *nU* = 2,270) | | |
| Sepsis | | |
| Yes (*nE* = 153; *nU* = 154) | 92.2 | 89.0 |
| No (*nE* = 2,119; *nU* = 2,116) | 90.0 | 88.7 |
| Heart failure | | |
| Yes (*nE* = 388; *nU* = 388) | 92.5 | 90.5 |
| No (*nE* = 1,884; *nU* = 1,882) | 89.7 | 88.3 |
| Stroke | | |
| Yes (*nE* = 203[a]; *nU* = 203[a]) | 80.8 | 83.3 |
| No (*nE* = 2,069; *nU* = 2,067) | 91.1 | 89.2 |
| Hygiene—Toileting (*nE* = 1,543[a]; *nU* = 1,541[a])[b] | | |
| Independent (*nE* = 72; *nU* = 72) | 100 | 98.6 |
| Setup or clean-up assistance (*nE* = 79; *nU* = 79) | 96.2 | 94.9 |
| Supervision or touching assistance (*nE* = 324; *nU* = 322) | 92.9 | 92.6 |
| Partial/moderate assistance (*nE* = 369; *nU* = 367) | 92.1 | 91.8 |
| Substantial/maximal assistance (*nE* = 341; *nU* = 342) | 93.6 | 90.6 |
| Dependent (*nE* = 358; *nU* = 359) | 86.6 | 86.9 |
| Mobility—Transfer from lying to sitting (*nE* = 1,895[a]; *nU* = 1,892[a]) | | |
| Independent (*nE* = 193; *nU* = 192) | 94.8 | 93.8 |
| Setup or clean-up assistance (*nE* = 113; *nU* = 113) | 86.7 | 88.5 |
| Supervision or touching assistance (*nE* = 531; *nU* = 530) | 93.6 | 92.3 |
| Partial/moderate assistance (*nE* = 622; *nU* = 621) | 90.5 | 88.9 |
| Substantial/maximal assistance (*nE* = 297; *nU* = 297) | 86.9 | 83.5 |
| Dependent (*nE* = 139; *nU* = 139) | 82.0 | 82.7 |

NOTE: Because of differences in sample sizes for the Expression data element and Understanding data element, we report sample sizes for each (*nE* = Expression; *nU* = Understanding). -

[a] Significant ($p < 0.05$) associations with expression and understanding as indicated by chi-square tests of independence.

[b] Toileting data not available for HHA patients.

Based on the research literature, we generated several hypotheses or expectations for associations between the data elements and the patient/resident characteristics. We expected the Expression and Understanding data elements to be related to stroke, toileting, and ability to transfer from lying to sitting, such that more-impaired patients/residents would be more likely to have suffered a stroke[50] and have less independence in ADLs.[51]

Overall, there were significant associations between age, length of stay, disposition at discharge, stroke, and both ADLs—toileting and ability to transfer from lying to sitting—and the Expression data elements (i.e., a4 and a6 in Tables 5.1 and 5.2) and Understanding data elements (i.e., a5 and a7 in Tables 5.1 and 5.2). There were no other significant associations. We review the statistical associations between variables next.

*Gender and Age*

- Overall, gender was not associated with the Expression or Understanding data elements. For both males and females, 90 percent were able to express themselves without difficulty, and 88 percent were able to understand others without difficulty. Similarly, gender was not associated with patient/resident ability to express themselves or understand others in any of the four settings. *This is consistent with our expectations that gender is not related to performance on these data elements.*
- There were significant associations overall between age and Expression data elements $\left(\chi^2_{(4)} = 24.34, p < 0.001\right)$ and age and Understanding data elements $\left(\chi^2_{(4)} = 43.33, p < 0.001\right)$. Although most patients/residents were able to express themselves and understand others without difficulty, 12 percent of those ages 18–44, 12 percent of those ages 75–89, and 13 percent of those 90 or older had difficulty with expression, and 12 percent of those ages 18–44, 13 percent of those ages 75–89, and 20 percent of those 90 or older had difficulty understanding others. Similar patterns significantly emerged in HHAs and SNFs, where difficulty with expression and understanding were more prevalent among the youngest (18–44) and oldest (75–89, 90 and older) age groups. In LTCHs, age was significantly associated only with Expression: The youngest (18–44) and oldest (75–89, 90 and older) age groups had greater difficulty with expression. Age was not associated with the Expression or Understanding data elements in the IRF setting. *As anticipated, this suggests that the Expression and Understanding data elements are capturing communication or cognitive impairments that are more prevalent in older patients/residents.*

*Length of Stay and Disposition at Discharge*

- Length of stay was also associated with Expression data elements $\left(F_{(1,2596)} = 20.40, p < 0.001\right)$ and Understanding data elements overall $\left(F_{(1,2595)} = 40.40, p < 0.001\right)$. On average, length of stay was longer for those having difficulty with

[50] Borthwick, 2012.

[51] Horowitz, 1994

expression: 25.2 days (SD = 15.0) compared with 21.2 days (SD = 12.4) for those without difficulty. Similarly, length of stay was longer for those with difficulty understanding others: 26.2 days (SD = 15.6) compared with 21.0 days on average (SD = 12.3) for those without difficulty. The same significant association between the Expression data elements and length of stay was found in all settings except SNFs. The significant association between Understanding data elements and length of stay was found in all four settings. *We did not anticipate this finding. However, this association is consistent with the idea that patients/residents who have difficulty expressing themselves and understanding others may have poorer health status overall or associated clinical conditions (e.g., traumatic brain injury) that require more time receiving PAC services*.

- Disposition at discharge was associated with the Expression data elements overall $\left(\chi^2_{(7)} = 46.01, p < 0.001\right)$: 5–10 percent of those discharged to an HHA, IRF, hospital, or home had difficulty with expression, whereas 14–23 percent of those discharged to an SNF, hospice, LTCH, or other location had difficulty with expression. Specifically, only 5 percent of those discharged to an HHA had difficulty with expression, while 23 percent of those discharged to an LTCH had difficulty with expression. For all four settings, disposition at discharge was significantly associated with expression. For HHAs, those discharged to an HHA, IRF, or LTCH were more likely to have difficulty with expression than in all other placements. For IRFs and LTCHs, those discharged to an SNF, "other" setting, or hospice were more likely to have difficulty with expression than in all other placements. Finally, for SNFs, those discharged to an LTCH, hospital, or "other" setting were more likely to have difficulty with expression than in all other placements.
- The Understanding data elements were similarly associated with disposition at discharge overall $\left(\chi^2_{(7)} = 32.09, p < 0.001\right)$: For discharges to most settings (i.e., home, hospital, HHA, IRF, LTCH, SNF), at most, 13 percent of patients/residents had difficulty understanding. However, 18 percent of patients/residents discharged to an "other" setting and 20 percent of patients/residents discharged to hospice had difficulty understanding. Disposition at discharge was associated with the Understanding data elements in only the SNF setting, such that those discharged to a hospital or "other" setting were more likely to have difficulty with understanding than in all other placements. *We did not anticipate these findings. However, these associations are consistent with the explanation proposed for longer length of stay: Patients/residents who have difficulty expressing themselves or communicating with others may have poorer health status overall or associated clinical conditions (e.g., traumatic brain injury) that require more time receiving care and/or more-intensive types of health care services.*

*Clinical Conditions*

- Frequencies showed no association between sepsis and difficulty with expression (6 percent with and 10 percent without had difficulty with expression) or understanding (9 percent with and 11 percent without had difficulty understanding). Likewise, there were no associations between sepsis and expression or understanding in any of the four settings. *This is consistent with our expectations, as sepsis is not described in*

*the literature as a common correlate or cause of problems with expression and understanding.*

- Patients/residents with and without heart failure also had similar rates of difficulty with expression (7 percent and 10 percent for those with and without heart failure, respectively) and understanding (10 percent and 12 percent for those with and without heart failure, respectively). Likewise, there were no associations between heart failure and expression or understanding in any of the four settings. *This is consistent with our expectations, for the same reasons as the lack of association with sepsis.*
- There was, however, an association between stroke and both the Expression data elements $\left(\chi^2_{(1)} = 22.91, p < 0.001\right)$ and the Understanding data elements $\left(\chi^2_{(1)} = 6.98, p < 0.01\right)$: Stroke patients/residents were more likely to have difficulty with both expression and understanding (19 percent versus 9 percent for expression and 17 percent versus 11 percent for understanding for patients/residents with and without stroke, respectively). The association between expression and stroke was also evident among patients/residents in all settings except HHAs, whereas the association between understanding and stroke was found in only the LTCH and SNF settings. *We anticipated associations between stroke and these data elements. These findings support the validity of the Expression and Understanding data elements, in that they were able to differentiate a group of patients/residents who are clinically much more likely to experience communication impairments.*

*ADLs: Toileting and Ability to Transfer from Lying to Sitting*

- The Expression and Understanding data elements were also associated (as expected) with independence levels on toileting (Expression: $\chi^2_{(5)} = 23.57, p < 0.01$; Understanding: $\chi^2_{(5)} = 15.29, p < 0.01$) and ability to transfer from lying to sitting (Expression: $\chi^2_{(5)} = 27.88, p < 0.01$; Understanding: $\chi^2_{(5)} = 24.99, p < 0.01$), such that rates of ability to express and understand without difficulty tended to be higher for patients/residents with higher levels of independence on these ADLs. For example, 100 percent and 94.8 percent of patients/residents who were rated as independent on toileting and ability to transfer from lying to sitting, respectively, were able to express ideas without difficulty, whereas rates for ability to express ideas without difficulty were much lower for those rated as dependent on these two ADLs (86.6 percent for toileting and 82.0 percent for ability to transfer from lying to sitting). Similarly, 98.6 percent and 93.8 percent of patients/residents who were rated as independent on toileting and transfer from lying to sitting, respectively, were able to understand, whereas rates for ability to understand were lower for those rated as dependent on these two ADLs (86.9 percent for toileting and 82.7 percent for ability to transfer from lying to sitting).This trend was also observed for toileting among LTCH patients (Expression: $\chi^2_{(5)} = 15.69, p < 0.01$; Understanding: $\chi^2_{(5)} = 13.67, p < 0.05$) and among SNF residents for Understanding only $\left(\chi^2_{(5)} = 11.61, p < 0.05\right)$. Furthermore, this trend was observed for ability to transfer from lying to sitting among HHA (Expression: $\chi^2_{(5)} = 20.24, p < 0.01$; Understanding: $\chi^2_{(5)} = 20.53, p < 0.01$), IRF (Expression: $\chi^2_{(5)} = 17.17, p < 0.01$;

Understanding: $\chi^2_{(5)} = 14.88, p < 0.05$), and LTCH patients (Expression: $\chi^2_{(5)} = 17.53, p < 0.01$; Understanding: $\chi^2_{(5)} = 12.16, p < 0.05$), but not for SNF residents. *These findings were consistent with our expectation and likely reflect the greater impairment overall of patients/residents with communication impairments.*

### Time to Complete

Table 5.4 shows the time, on average, to complete the Expression and Understanding data elements overall and by setting for both versions. On average, the three–data element version took 0.8 minutes (SD = 0.4) to complete, ranging from 0.8 minutes (SD = 0.4) in HHAs and IRFs to 0.9 minutes (SD = 0.4) in LTCHs and SNFs. The two–data element version, on average, took 0.7 minutes (SD = 0.3) to complete, ranging from 0.7 minutes (SD = 0.3–0.4) in IRFs, LTCHs, and SNFs to 0.8 minutes (SD = 0.3) in HHAs. For the three–data element version, there was a significant association between setting and time to complete $\left(F_{(3,811)} = 5.01, p < 0.01\right)$: It took significantly more time to complete in LTCHs than in HHAs $\left(t_{(811)} = 3.19, p < 0.01\right)$. Likewise, for the two–data element version, there was a significant association between setting and time to complete $\left(F_{(3,833)} = 4.13, p < 0.01\right)$: It took significantly more time to complete in HHAs than in IRFs $\left(t_{(833)} = 3.40, p < 0.001\right)$. There were no other significant differences among settings for either version.

**Table 5.4. Time to Complete for Expression and Understanding Data Elements (minutes)**

| Version | Characteristic | HHA (A: *n* = 145; B: *n* = 251) | IRF (A: *n* = 330; B: *n* = 169) | LTCH (A: *n* = 136; B: *n* = 165) | SNF (A: *n* = 204; B: *n* = 252) | Overall (A: *n* = 815; B: *n* = 837) |
|---|---|---|---|---|---|---|
| Market Group A (three–data element: a3, a4, a5) | Mean (SD) | 0.8 (0.4)[a] | 0.8 (0.4) | 0.9 (0.4)[a] | 0.9 (0.4) | 0.8 (0.4) |
| Market Group B (two–data element: a6 and a7) | Mean (SD) | 0.8 (0.3)[a] | 0.7 (0.4) | 0.7 (0.3)[a] | 0.7 (0.3) | 0.7 (0.3) |

[a] Significant ($p < 0.01$) differences between settings as determined by independent samples t-tests.

Time to complete was also evaluated according to urbanicity (urban versus nonurban), geographic region (Northeast, South, Midwest, West), facility ownership (for-profit versus nonprofit), and facility size (above or below setting-type median) to evaluate the generalizability of these performance results (see Tables A.29–A.32 in the appendix). In these sensitivity analyses, significant differences were found for time to complete the Expression and Understanding data elements in Market Group B (two–data element version) only. Specifically, assessors in the West region took significantly longer than assessors in the South region, but with a small effect size (0.30). Also, assessors in smaller facilities took longer than assessors in larger facilities, with a small to moderate effect size (0.40). There were no other significant differences

for time to complete the Expression and Understanding data elements in these sensitivity analyses.

### *Interrater Reliability*

Tables 5.5 and 5.6 show kappa interrater reliability coefficients for the three– and two–data element versions overall and by setting. As described in more detail in Volume 3, paired assessment data for interrater reliability evaluation were collected on a subset of the National Beta Test admission sample of patients/residents according to setting-level target totals. For example, each participating LTCH was asked to conduct 20 paired assessments to contribute to interrater reliability. Inclusion in interrater reliability data collection depended on paired facility staff and research nurse assessors' ability to schedule assessments. Results are presented for 959 patients/residents: 441 for the three–data element version and 518 for the two–data element version.

**Table 5.5. Interrater Reliability Kappa or Weighted Kappa for Expression and Understanding Data Elements: Market Group A**

| Data Element | HHA (*n* = 57) | IRF (*n* = 159) | LTCH (*n* = 112) | SNF (*n* = 113) | Overall (*n* = 441) |
|---|---|---|---|---|---|
| Speech clarity (a3) | 0.46 | — | 0.75 | — | — |
| Ability to express ideas and wants (a4) | 0.50 | — | — | — | 0.64 |
| Understanding verbal content (a5) | 0.58 | — | 0.60 | 0.43 | 0.59 |

NOTES: Interrater reliability not shown for data elements with proportions out of range for stable kappa estimate (per study power calculations). Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

**Table 5.6. Interrater Reliability Kappa or Weighted Kappa for Expression and Understanding Data Elements: Market Group B**

| Data Element | HHA (*n* = 140) | IRF (*n* = 99) | LTCH (*n* = 125) | SNF (*n* = 154) | Overall (*n* = 518) |
|---|---|---|---|---|---|
| Expresses ideas and wants (a6) | 0.58 | 0.38 | 0.43 | 0.26 | 0.42 |
| Understands verbal content (a7) | 0.39 | 0.32 | 0.22 | 0.36 | 0.32 |

NOTE: Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

For the three–data element version overall, kappa was good (0.64) for "ability to express ideas and wants" and moderate (0.59) for "understanding verbal content." For the three–data element version at the setting level, kappas were moderate to good and ranged from 0.46 to 0.58 in HHAs, 0.60 to 0.75 in LTCHs, and 0.43 for "understands verbal content" in SNFs. Kappas are not presented for IRFs because prevalence rates were out of range for stable estimates. For the two–data element version overall, kappa was moderate (0.42) for "expresses ideas and wants" and fair (0.32) for "understands verbal content." For the two–data element version at the setting

level, kappas were fair to moderate and ranged from 0.39 to 0.58 in HHAs, 0.32 to 0.38 in IRFs, 0.22 to 0.43 in LTCHs, and 0.26 to 0.36 in SNFs. Generally speaking, kappas overall and at the setting level tended to be higher for the three–data element version than the two–data element version.

Interrater reliability (kappa) was also evaluated according to urbanicity (urban versus nonurban), geographic region (Northeast, South, Midwest, West), facility ownership (for-profit versus nonprofit), and facility size (above or below setting-type median) to evaluate the generalizability of these performance results (see Tables A.33−A.40 in the appendix). No noteworthy differences were found for interrater reliability of the Expression and Understanding data elements in these sensitivity analyses.

Tables 5.7 and 5.8 show the percent agreement for all data elements overall and by setting. Overall percent agreement for the three–data element version was high for all data elements, ranging from 93 percent to 95 percent. Overall percent agreement was slightly lower for the two–data element version, with 89 percent and 86 percent for expression and understanding, respectively. For the three–data element version at the setting level, percent agreement ranged from 81 percent to 93 percent in HHAs, 94 percent to 97 percent in IRFs, 94 percent to 95 percent in LTCHs, and 93 percent to 95 percent in SNFs. For the two–data element version, percent agreements for understanding and expression were 87 percent and 93 percent in HHAs, 86 percent in IRFs, 82 percent and 86 percent in LTCHs, and 90 percent and 91 percent in SNFs, respectively. Overall, kappa coefficients and percent agreement were better in the three–data element version. Similarly, percent agreement at the setting level tended to be higher for the three–data element version.

**Table 5.7. Interrater Reliability—Percent Agreement for Expression and Understanding Data Elements: Market Group A**

| Data Element | HHA (*n* = 57) | IRF (*n* = 159) | LTCH (*n* = 112) | SNF (*n* = 113) | Overall (*n* = 441) |
|---|---|---|---|---|---|
| Speech clarity (a3) | 93 | 94 | 95 | 96 | 95 |
| Ability to express ideas and wants (a4) | 81 | 97 | 94 | 95 | 93 |
| Understanding verbal content (a5) | 84 | 95 | 94 | 92 | 93 |

**Table 5.8. Interrater Reliability—Percent Agreement for Expression and Understanding Data Elements: Market Group B**

| Data Element | HHA (*n* = 140) | IRF (*n* = 99) | LTCH (*n* = 125) | SNF (*n* = 154) | Overall (*n* = 518) |
|---|---|---|---|---|---|
| Expresses ideas and wants (a6) | 93 | 86 | 86 | 91 | 89 |
| Understands verbal content (a7) | 87 | 86 | 82 | 90 | 86 |

### *Day 3, 5, 7 Repeat Assessment Evaluation*

Tables 5.9 and 5.10 summarizes patterns of change for Expression and Understanding data elements across the repeat assessment days. Patterns are characterized as "no change" (scores stay the same across assessment days), "improve" (steady improvement across assessment days), "worsen" (steady decline across assessment days), and "fluctuate" (scores go up and down across assessment days).

**Table 5.9. Day 3, 5, and 7 Repeat Assessment Results for Expression and Understanding Data Elements: Market Group A (percent)**

| Data Element | HHA (*n* = 16) | IRF (*n* = 92) | LTCH (*n* = 45) | SNF (*n* = 92) | Overall (*n* = 245) |
|---|---|---|---|---|---|
| Speech clarity (a3) | | | | | |
| No change | 88 | 92 | 89 | 99 | 94 |
| Worsen | 0 | 2 | 9 | 0 | 2 |
| Improve | 13 | 5 | 2 | 1 | 4 |
| Fluctuate | 0 | 0 | 0 | 0 | 0 |
| Ability to express ideas and wants (a4) | | | | | |
| No change | 88 | 91 | 93 | 92 | 92 |
| Worsen | 0 | 5 | 7 | 2 | 4 |
| Improve | 13 | 2 | 0 | 4 | 3 |
| Fluctuate | 0 | 1 | 0 | 1 | 1 |
| Understanding verbal content (a5) | | | | | |
| No change | 81 | 87 | 93 | 80 | 85 |
| Worsen | 6 | 8 | 7 | 8 | 7 |
| Improve | 13 | 3 | 0 | 5 | 4 |
| Fluctuate | 0 | 2 | 0 | 7 | 3 |

**Table 5.10. Day 3, 5, and 7 Repeat Assessment Results for Expression and Understanding Data Elements: Market Group B (percent)**

| Data Element | HHA (*n* = 79) | IRF (*n* = 30) | LTCH (*n* = 26) | SNF (*n* = 100) | Overall (*n* = 235) |
|---|---|---|---|---|---|
| Expresses ideas and wants (a6) | | | | | |
| No change | 90 | 83 | 88 | 89 | 89 |
| Worsen | 5 | 0 | 12 | 4 | 5 |
| Improve | 3 | 10 | 0 | 5 | 4 |
| Fluctuate | 3 | 7 | 0 | 2 | 3 |
| Understands verbal content (a7) | | | | | |
| No change | 90 | 73 | 84 | 84 | 84 |
| Worsen | 5 | 7 | 12 | 6 | 6 |
| Improve | 1 | 10 | 4 | 5 | 4 |
| Fluctuate | 4 | 10 | 0 | 5 | 5 |

As described in more detail in Volume 3, repeat assessment data were collected on a subset of the National Beta Test admission sample of patients/residents according to setting-level target totals. For example, each participating HHA was asked to contribute five patients for repeat assessment. Inclusion in repeat assessment data collection depended on assessor ability to complete the initial admission assessment on Day 3, availability of assessor to return for repeat assessments on Days 5 and 7, and willingness of patient/resident to complete multiple assessments. Results are presented for 480 patients/residents: 245 for the three–data element version and 235 for the two–data element version. For both versions overall, there were no statistically significant differences on any of the data elements across assessment days. Responses to the data elements were similar for both versions across assessment days and showed very little change, ranging from 85 percent to 94 percent (three–data element version) and 84 percent to 89 percent (two–data element version). Specifically, Speech Clarity had the highest consistency (94 percent). For both the three– and two–data element versions, "expression" data elements remained relatively stable at 92 percent and 89 percent, respectively, as did the "understanding" data elements at 85 percent and 84 percent, respectively.

### *Admission to Discharge*

Tables 5.11 and 5.12 summarize patterns of change from admission to discharge for both versions of the Expression and Understanding data elements. Patterns are characterized as "no change" (scores stay the same at admission and discharge), "improve" (scores reflect improvement from admission to discharge), and "worsen" (scores indicate a decline from admission to discharge). As described in more detail in Volume 3, discharge data were collected on a subset of the National Beta Test admission sample of patients/residents. Availability of discharge data depended on advance notification of discharge and the ability to schedule

assessments among the facility staff assessors at each participating site. Results are based on data from 791 patients/residents: 465 of whom received the three–data element version and 326 of whom received the two–data element version. Responses from admission to discharge were consistent for the majority of data elements, with an overall range of 90 percent to 97 percent experiencing no change. Similar to Day 3, 5, 7 results, there were no overall statistical differences in scores between admission and discharge for either the two– or three–data element version.

**Table 5.11. Admission to Discharge Results for Expression and Understanding Data Elements: Market Group A (percent)**

| Data Element | HHA (*n* = 53) | IRF (*n* = 226) | LTCH (*n* = 42) | SNF (*n* = 144) | Overall (*n* = 465) |
|---|---|---|---|---|---|
| Speech clarity (a3) | | | | | |
| No change | 100 | 96 | 89 | 99 | 97 |
| Worsen | 0 | 1 | 9 | 1 | 1 |
| Improve | 0 | 2 | 2 | 0 | 2 |
| Ability to express ideas and wants (a4) | | | | | |
| No change | 85 | 92 | 98 | 96 | 93 |
| Worsen | 9 | 4 | 0 | 3 | 4 |
| Improve | 6 | 4 | 2 | 1 | 3 |
| Understanding verbal content (a5) | | | | | |
| No change | 85 | 89 | 100 | 96 | 91 |
| Worsen | 4 | 6 | 0 | 3 | 4 |
| Improve | 11 | 6 | 0 | 1 | 5 |

**Table 5.12. Admission to Discharge Results for Expression and Understanding Data Elements: Market Group B (percent)**

| Data Element | HHA (*n* = 91) | IRF (*n* = 112) | LTCH (*n* = 41) | SNF (*n* = 82) | Overall (*n* = 326) |
|---|---|---|---|---|---|
| Expresses ideas and wants (a6) | | | | | |
| No change | 93 | 90 | 93 | 94 | 92 |
| Worsen | 1 | 6 | 5 | 4 | 4 |
| Improve | 6 | 4 | 2 | 2 | 4 |
| Understands verbal content (a7) | | | | | |
| No change | 88 | 89 | 95 | 90 | 90 |
| Worsen | 3 | 6 | 5 | 5 | 5 |
| Improve | 9 | 5 | 0 | 5 | 5 |

### *Assessor Feedback*

Research nurses reported that the strength of the Expression and Understanding data elements is that they provide consistent information across PAC settings and facilities. The research nurses further noted that having consistent information during patient/resident transfer is particularly clinically relevant, perhaps because transfers are a time when staff may not yet know a patient's/resident's communication abilities through experience with that patient/resident but still need to communicate with the patient/resident during the initial intake. In the assessor survey, the Expression and Understanding data elements were rated among the top five data elements in terms of clinical utility regardless of which version was used. In addition, facility/agency staff and research nurses considered these data elements to be of low burden.

In focus groups, research nurses stated that both the two– and three–data element versions of these data elements are easy to collect. Some also noted that not having to adhere to a script helped the assessors connect with the patient/resident and allowed for assessment through more-natural communication and observation.

## Summary

Results for Expression and Understanding indicate high overall support for cross-setting standardization. Expression and Understanding were ranked very highly in terms of clinical utility by assessors. The three–data element set had slightly higher percent agreement and interrater reliability than the two–data element version, with moderate to good reliability. Results of the Day 3, 5, and 7 repeat assessment evaluation indicate that both the three– and two–data element Expression and Understanding data elements were fairly stable regardless of the day of assessment; that is, the day the assessment was administered did not affect data element performance. Both data element sets were also fairly stable from admission to discharge. In addition, the associations between Expression and Understanding and patient/resident characteristics somewhat aligned with our expected results, indicating validity of the data elements. In the National Beta Test, both the three–data element version and the two–data element version of Expression and Understanding showed high feasibility and low burden based on the time to complete and feedback from the assessors.

# 6. Behavioral Signs and Symptoms

## Data Element Description

The Behavioral Signs and Symptoms data elements assess the presence and frequency of behavioral symptoms, the impact of behavioral symptoms on the patient/resident and others, including risk for physical injury, interference with patient/resident care, interference with patient's/resident's participation in activities, intrusion on privacy of others, the disruption of delivery of care or living environment of others, and the presence and frequency of rejection of care. Patients/residents with these behavioral symptoms may require more case management time because of a need for supervision and/or therapies addressed at addressing behaviors,[52] may have poorer quality of life and interpersonal relationships, and may be at risk for injury, isolation, and inactivity.[53] These behaviors can also disrupt the institutional or home environment and affect the safety and privacy of other patients/residents and caregivers.[54] Assessment and documentation of these behavioral symptoms can help inform care planning, staffing, and patient/resident transitions.

The Behavioral Signs and Symptoms data elements are assessed by reviewing the medical record, interviewing staff and others who interact with the patient/resident, and observing the patient/resident. Behavioral symptoms are assessed by the OASIS and the MDS; the version of the data elements tested in the National Beta Test was derived from the MDS 3.0. In the National Beta Test, the data elements were included in the repeat assessment evaluation and therefore were evaluated repeatedly on the same patient/resident by the same assessor on Days 3, 5, and 7. The Behavioral Signs and Symptoms data elements are shown in Figures 6.1–6.3.

[52] Diwan and Phillips, 2001.

[53] Voyer et al., 2005.

[54] Voyer et al., 2005; Tan, Wong, and Allen, 2005; Kaufer et al., 1998.

**Figure 6.1. Behavioral Signs and Symptoms Presence and Frequency**

**INSTRUCTIONS: ALL ITEMS IN MODULE H: BEHAVIORAL SIGNS AND SYMPTOMS ARE BASED ON STAFF/CAREGIVER INPUT OR CHART REVIEW. DO NOT ASK PATIENT/RESIDENT.**

**RECORD RESPONSES BASED ON BEHAVIORS IN THE PAST 3 DAYS.**

**H1a.** Physical behavioral symptoms directed toward others

(e.g., hitting, kicking, pushing, scratching, grabbing, abusing others sexually)

- ☐ 0 = Behavior not exhibited
- ☐ 1 = Behavior of this type occurred 1 day
- ☐ 2 = Behavior of this type occurred 2 days, but less than daily
- ☐ 3 = Behavior of this type occurred daily
- ☐ 9 = **Unknown or unable to assess**

**H1b.** Verbal behavioral symptoms directed toward others

(e.g., threatening others, screaming at others, cursing at others)

- ☐ 0 = Behavior not exhibited
- ☐ 1 = Behavior of this type occurred 1 day
- ☐ 2 = Behavior of this type occurred 2 days, but less than daily
- ☐ 3 = Behavior of this type occurred daily
- ☐ 9 = **Unknown or unable to assess**

**H1c.** Other behavioral symptoms not directed toward others

(e.g., physical symptoms such as hitting or scratching self, pacing, rummaging, public sexual acts, disrobing in public, throwing or smearing food or bodily wastes, or verbal/vocal symptoms like screaming, disruptive sounds)

- ☐ 0 = Behavior not exhibited
- ☐ 1 = Behavior of this type occurred 1 day
- ☐ 2 = Behavior of this type occurred 2 days, but less than daily
- ☐ 3 = Behavior of this type occurred daily
- ☐ 9 = **Unknown or unable to assess**

**Figure 6.2. Behavioral Signs and Symptoms Impact on Patient/Resident**

| |
|---|
| **IF ALL RESPONSES TO H1a, H1b, AND H1c ARE CODED AS EITHER "(0) – BEHAVIOR NOT EXHIBITED") OR "(9) – UNKNOWN OR UNABLE TO ASSESS," SKIP TO H4** |
| **IMPACT ON PATIENT/RESIDENT**<br><br>**INSTRUCTIONS: CONSIDERING ALL THE BEHAVIORAL SYMPTOMS NOTED IN H1a, H1b, AND H1c, DID ANY OF THE IDENTIFIED SYMPTOM(S):** |
| **H2a.** Put the patient/resident at significant risk for physical illness or injury?<br><br>☐ 0 = No<br>☐ 1 = Yes<br>☐ 9 = **Unknown or unable to assess** |
| **H2b.** Significantly interfere with the patient's/resident's care?<br><br>☐ 0 = No<br>☐ 1 = Yes<br>☐ 9 = **Unknown or unable to assess** |
| **H2c.** Significantly interfere with the patient's/resident's participation in activities or social interaction?<br><br>☐ 0 = No<br>☐ 1 = Yes<br>☐ 8 = Not Applicable<br>☐ 9 = **Unknown or unable to assess** |

**Figure 6.3. Behavioral Signs and Symptoms Impact on Others**

**IMPACT ON OTHERS**

**INSTRUCTIONS: CONSIDERING ALL THE BEHAVIORAL SYMPTOMS NOTED IN H1a, H1b, AND H1c, DID ANY OF THE IDENTIFIED SYMPTOM(S):**

**H3a.** Put others at significant risk for physical injury?

- ☐ 0 = No
- ☐ 1 = Yes
- ☐ 9 = **Unknown or unable to assess**

**H3b.** Significantly intrude on the privacy or activity of others?

- ☐ 0 = No
- ☐ 1 = Yes
- ☐ 9 = **Unknown or unable to assess**

**H3c.** Significantly disrupt the delivery of care or living environment of others?

- ☐ 0 = No
- ☐ 1 = Yes
- ☐ 9 = **Unknown or unable to assess**

**H4.** Did the patient/resident reject evaluation or care (e.g., bloodwork, taking medications, ADL assistance) that is offered by members of the care team or caregiver and necessary to achieve the patient's/resident's goals for health and well-being?

**Do not include behaviors that have already been addressed (e.g., by discussion or care planning with the patient/resident or family), and determined to be consistent with patient/resident values, preferences, or goals.**

- ☐ 0 = Behavior not exhibited
- ☐ 1 = Behavior of this type occurred 1 day
- ☐ 2 = Behavior of this type occurred 2 days, but less than daily
- ☐ 3 = Behavior of this type occurred daily
- ☐ 9 = **Unknown or unable to assess**

## Testing Objectives

Across all National Beta Test assessments of the Behavioral Signs and Symptoms data elements, it was rare for assessors to indicate that patients/residents exhibited any of these behavioral symptoms, and the higher frequencies of occurrence (two out of three days or daily) were not observed often enough to provide sufficient data for most analyses. For this reason (and to enable evaluation of these data elements), all four-category data element responses (i.e., for data elements H1a, H1b, H1c, and H4 in Figures 6.1–6.3) were dichotomized for analysis. Basic descriptive statistics (e.g., frequencies) are presented for admission data. We also examined whether exhibiting any behavioral symptom at admission (physical, verbal, or other) was associated with patient/resident characteristics and clinical conditions. For admission data, feasibility (rates of missingness and time to complete) and interrater reliability (kappa and percent agreement) were examined. The Behavioral Signs and Symptoms data elements were also administered repeatedly on the same patient/resident by the same assessor to test the effects of conducting the admission assessment on Days 3, 5, and 7. Therefore, an additional objective was to understand whether there were significant and meaningful differences in rates or scores depending on the day a patient/resident was assessed. Lastly, frequencies at admission and discharge were compared to inform stability or possible change over time.

## Results

### *Feasibility*

#### Frequencies/Missing

Table 6.1 shows the frequency percentages and counts for the Behavioral Signs and Symptoms data elements (physical, verbal, and other symptoms, and rejection of care) overall and by setting. Frequency data for the Behavioral Signs and Symptoms data elements were collected on 2,954 patients/residents: 625 in HHAs, 772 in IRFs, 471 in LTCHs, and 1,086 in SNFs. Overall, more than 94 percent of the sample were administered the Behavioral Signs and Symptoms data elements. Among these, missing data at the data element level ranged from 0.2 percent to 0.5 percent, with minimal setting differences. As expected, prevalence rates for all Behavioral Signs and Symptoms data elements were extremely low. Rejection of evaluation and care, the most frequently observed behavior, was exhibited by only 5.5 percent of all patients/residents, with a range from 3.8 percent in HHAs to 8.3 percent in SNFs. Overall, less than 1 percent of patients/residents exhibited physical behavioral symptoms directed toward others, with the greatest prevalence in LTCHs (1.5 percent). Only 1.3 percent of total patients/residents exhibited verbal behavioral symptoms directed toward others, with the greatest prevalence (1.9 percent) observed in LTCHs. Less than 1 percent of all patients/residents

exhibited other behavioral symptoms directed toward others, with the greatest prevalence (1 percent) in SNFs. A total of 59 patients/residents exhibited any behavioral symptoms (including physical, verbal, or other). For these 59 patients/residents, these symptoms most commonly interfered with the patients'/residents' care or participation, and less frequently put the patient/resident at risk, put others at risk, interfered with others, intruded on the privacy of others, or disrupted others' care (see Table 6.1 for overall and setting-specific results).

**Table 6.1. Overall and Setting-Specific Response Frequencies for Behavioral Signs and Symptoms Data Elements (percent)**

| Data Element | HHA (*n* = 625) | IRF (*n* = 772) | LTCH (*n* = 471) | SNF (*n* = 1,086) | Overall (*n* = 2,954) |
|---|---|---|---|---|---|
| Physical behavioral symptoms directed toward others (h1a)—behavior exhibited (count) | 0.3 (2) | 0.4 (3) | 1.5 (7) | 0.2 (2) | 0.5 (14) |
| Verbal behavioral symptoms directed toward others (h1b)—Behavior exhibited (count) | 1.3 (8) | 1.4 (11) | 1.9 (9) | 1.0 (11) | 1.3 (39) |
| Other behavioral symptoms directed toward others (h1c)—Behavior exhibited (count) | 0.3 (2) | 0.7 (5) | 0.4 (2) | 1.0 (11) | 0.7 (20) |
| Did patient/resident reject evaluation or care (h4)—Behavior exhibited (count) | 3.8 (24) | 8.3 (64) | 6.2 (29) | 4.2 (46) | 5.5 (163) |
| Percentage of patients/residents exhibiting any symptoms (physical, verbal, other) (count)[a] | 1.4 (9) | 2.1 (16) | 3.0 (14) | 1.8 (20) | 2.0 (59) |
| Did behavioral symptoms put patient/resident at risk (h2a)—Yes (count) | 0 | 6.3 (1) | 21.4 (3) | 5.0 (1) | 8.5 (5) |
| Did behavioral symptoms interfere with care (h2b)—Yes (count) | 11.1 (1) | 18.8 (3) | 28.6 (4) | 5.0 (1) | 15.3 (9) |
| Did behavioral symptoms interfere with patient/resident participation (h2c)—Yes (count) | 44.4 (4) | 12.5 (2) | 28.6 (4) | 10.0 (2) | 20.3 (12) |
| Did behavioral symptoms put others at risk (h3a)—Yes (count) | 11.1 (1) | 6.3 (1) | 14.3 (2) | 5.0 (1) | 8.5 (5) |
| Did behavioral symptoms intrude on privacy of others (h3b)—Yes (count) | 11.1 (1) | 6.3 (1) | 14.3 (2) | 0 | 6.8 (4) |
| Did behavioral symptoms disrupt care for others (h3c)—Yes (count) | 22.2 (2) | 18.8 (3) | 7.1 (1) | 0 | 10.2 (6) |

NOTE: Because of low prevalence, frequencies are presented as percentages (with counts in parentheses).
[a] Percentages and counts for data elements h2a–h3c are in reference to sample sizes presented in this row.

### Known Groups Validity

Comparing the performance of patients/residents on the Behavioral Signs and Symptoms data elements with other patient/resident characteristics adds information about validity. If known or logical associations between patients/resident characteristics and data elements are observed in data from the National Beta test, this information contributes to the evidence that the data elements are valid, or assessing the construct that they are intended to capture.

Table 6.2 shows the frequencies for whether behavioral symptoms (physical, verbal, or other) were exhibited for the overall admission sample, stratified by patient/resident characteristics and

clinical groups as described in Chapter 1: gender (male or female, as documented by National Beta Test assessor), age (as categorized into the following ranges: 18–44, 45–64, 65–74, 75–89, 90 and over), length of stay (in days), disposition at discharge (e.g., to another PAC setting, home, to hospital), sepsis, heart failure, stroke, and two ADLs—toileting (not available for HHA patients) and ability to transfer from lying to sitting. As a reminder, these clinical conditions were chosen based on their common occurrence across settings, their frequent relationship with many of the data elements tested in the National Beta Test, and their availability in all four standardized assessments (OASIS, IRF-PAI, LCDS, MDS). Setting-specific results are presented in the appendix (Tables A.41–A.44). Because of the very low frequency of occurrence of these behaviors for the overall sample, we observe very few significant differences in rates of behavior occurrence based on these characteristics.

**Table 6.2. Frequencies for Any Type of Behavioral Symptom (Physical, Verbal, Other) Exhibited by Patient/Resident Characteristics and Clinical Groups (percent)**

| Patient/Resident Characteristics and Clinical Groups | Any Behavioral Symptom Exhibited (Yes) |
|---|---|
| Gender (*n* = 2,870) | |
| Male (*n* = 1,182) | 2.5 |
| Female (*n* = 1,688) | 1.7 |
| Age (*n* = 2,860) | |
| 18–44 (*n* = 38) | 5.3 |
| 45–64 (*n* = 304) | 2.6 |
| 65–74 (*n* = 903) | 1.4 |
| 75–89 (*n* = 1,311) | 1.9 |
| 90 and older (*n* = 304) | 3.0 |
| Length of stay (*n* = 2,537; mean, SD) | Yes: 21.4 (10.6)<br>No: 21.6 (12.8) |
| Disposition at discharge (*n* = 2,825) | |
| Home (*n* = 1,315) | 2.1 |
| Hospital (*n* = 194) | 1.0 |
| Hospice (*n* = 40) | 2.5 |
| SNF (*n* = 278) | 3.6 |
| IRF (*n* = 51) | 3.9 |
| HHA (*n* = 617) | 0.8 |
| LTCH (*n* = 12) | 8.3 |
| Other (*n* = 318) | 2.2 |

| Patient/Resident Characteristics and Clinical Groups | Any Behavioral Symptom Exhibited (Yes) |
|---|---|
| Clinical conditions (*n* = 2,204) | |
| Sepsis | |
| Yes (*n* = 148) | 1.4 |
| No (*n* = 2,056) | 2.0 |
| Heart failure | |
| Yes (*n* = 375) | 1.6 |
| No (*n* = 1,829) | 2.1 |
| Stroke | |
| Yes (*n* = 194) | 2.6 |
| No (*n* = 2,010) | 1.9 |
| Hygiene—Toileting (*n* = 1,495)[a] | |
| Independent (*n* = 72) | 2.8 |
| Setup or clean-up assistance (*n* = 77) | 0.0 |
| Supervision or touching assistance (*n* = 318) | 2.2 |
| Partial/moderate assistance (*n* = 357) | 1.4 |
| Substantial/maximal assistance (*n* = 330) | 3.9 |
| Dependent (*n* = 341) | 2.9 |
| Mobility—Transfer from lying to sitting (*n* = 1842) | |
| Independent (*n* = 192) | 1.0 |
| Setup or clean-up assistance (*n* = 108) | 2.8 |
| Supervision or touching assistance (*n* = 522) | 1.9 |
| Partial/moderate assistance (*n* = 609) | 1.6 |
| Substantial/maximal assistance (*n* = 281) | 4.3 |
| Dependent (*n* = 130) | 1.5 |

[a] Toileting data not available for HHA patients.

Although we did not have hypotheses or expectations for all characteristics and conditions listed in this table, we did expect the Behavioral Signs and Symptoms data elements to be related to length of stay, toileting, and ability to transfer from lying to sitting, such that more-impaired patients/residents would tend to have a longer length of stay[55] and have less independence in ADLs.[56]

Overall, none of the patient/resident characteristics or clinical groups were significantly associated with whether any behavioral symptoms were exhibited. At the setting level, with the exception of disposition at discharge in SNFs, age in HHAs, and stroke in LTCHs, there were no

[55] Beck, Rossby, and Baldwin, 1991.

[56] Beck et al., 1998.

significant associations with whether behavioral symptoms were exhibited. Next, we describe the results contained in Table 6.2 despite the lack of significant findings and include setting-level significant differences when they occur.

*Gender and Age*

- For gender, 2 percent of males and females exhibited behavioral symptoms.
- In regard to age, 5 percent of patients/residents ages 18–44, 3 percent of patients/residents ages 45–64, 1 percent of patients/residents ages 65–74, 2 percent of patients/residents ages 75–89, and 3 percent of patients/residents 90 or older exhibited behavioral symptoms. In HHAs, however, there was a significant association between age and exhibiting behavioral symptoms in that, compared with all other age groups, a greater proportion in the youngest group (18–44) exhibited behavioral symptoms. *This association with the younger age group in HHAs was not anticipated and is somewhat surprising, given the literature on dementia-associated behavioral symptoms. These younger PAC patients who exhibit behavioral symptoms while receiving care from HHAs may have medical conditions or histories that explain their behavioral symptoms, such as traumatic brain injury, mental health conditions, or intellectual or developmental disabilities. It is also possible that older patients with behavioral symptoms are less likely to receive HHA care (i.e., more likely to receive PAC services in a facility), making the rates in younger HHA patients seem relatively higher.*

*Length of Stay and Disposition at Discharge*

- Rates of behavioral symptoms were also unrelated to length of stay. The average length of stay was 21.4 days (SD = 10.6) for those exhibiting any behavior and 21.6 days (SD = 12.8) for those exhibiting no behavior, and this trend of similar lengths of stay regardless of occurrence of behaviors was also observed at the setting level for all four settings.
- Similarly, although there was some variability in rates of exhibiting behavioral symptoms according to disposition at discharge, these differences were not significant overall. Specifically, 8 percent of those discharged to LTCHs exhibited behavioral symptoms, whereas the rates of behaviors ranged from 1 percent to 4 percent for patients/residents discharged to all other settings (i.e., home, hospital, HHA, IRF, SNF, other). At the setting level, however, there was a significant association among SNF residents between disposition at discharge and whether behavioral symptoms were exhibited. Relative to discharges to all other placements, those discharged from an SNF to an LTCH or hospice were more likely to have exhibited behavioral symptoms. *This suggests that SNF residents with behavioral symptoms may be more seriously ill than SNF residents without behavioral symptoms.*

*Clinical Conditions*

- None of the clinical conditions were associated with behavioral symptoms in the sample overall. Frequencies show that 1 percent of those with sepsis and 2 percent without exhibited behavioral symptoms. Furthermore, 2 percent of those with heart

failure and 2 percent of those without exhibited behavioral symptoms, and 2 percent of stroke and 2 percent of nonstroke patients/residents exhibited behavioral symptoms. However, one association was observed at the setting level. In LTCH patients, stroke was significantly associated with exhibiting behavioral symptoms, such that stroke patients were more likely than nonstroke patients to have exhibited behavioral symptoms. *Behavioral and psychological symptoms secondary to stroke are not uncommon, including irritability and anger.*[57] *The fact that this association was present only in one setting may be related to the small numbers of patients/residents with behavior symptoms overall.*

### Time to Complete

Table 6.3 shows average time to complete Behavioral Signs and Symptoms data elements overall and by setting. Overall mean time to complete the Behavioral Signs and Symptoms data elements was 1.4 minutes (SD = 0.8 minutes). Setting-specific time to complete ranged from 1.4 minutes (SD = 0.7-0.8) in HHAs and IRFs to 1.5 minutes (SD = 0.7-0.8) in LTCHs and SNFs. There were no significant differences in time to complete among settings.

**Table 6.3. Time to Complete Behavioral Signs and Symptoms Data Elements (minutes)**

| Characteristic | HHA (*n* = 391) | IRF (*n* = 454) | LTCH (*n* = 283) | SNF (*n* = 415) | Overall (*n* = 1,543) |
|---|---|---|---|---|---|
| Mean (SD) | 1.4 (0.7) | 1.4 (0.8) | 1.5 (0.7) | 1.5 (0.8) | 1.4 (0.8) |

Time to complete was also evaluated according to urbanicity (urban versus nonurban), geographic region (Northeast, South, Midwest, West), facility ownership (for-profit versus nonprofit), and facility size (above or below setting-type median) to evaluate the generalizability of these performance results (see Tables A.45–A.48 in the appendix). Interestingly, significant differences were found in these sensitivity analyses for time to complete the Behavioral Signs and Symptoms data elements according to region. Northeast and West (both 1.6 minutes) are significantly higher than South (1.4 minutes) and Midwest (1.3 minutes), and nonurban is significantly higher than urban (1.7 minutes versus 1.4 minutes). However, effect sizes for these differences tended to be small (all were less than 0.3). There were no other significant differences in time to complete the Behavioral Signs and Symptoms data elements in these sensitivity analyses.

## *Interrater Reliability*

As described in more detail in Volume 3, paired assessment data for interrater reliability evaluation were collected on a subset of the National Beta Test admission sample of

[57] Cameron et al., 2006; Kim et al., 2002.

patients/residents according to setting-level target totals. For example, each participating LTCH was asked to conduct 20 paired assessments to contribute to interrater reliability. Inclusion in interrater reliability data collection depended on paired facility staff and research nurse assessors' ability to schedule these assessments. Kappa interrater reliability coefficients are not shown or discussed because, for all Behavioral Signs and Symptoms data elements, prevalence rates were extremely low and out of range for stable kappa estimates as determined by study power calculations. Similarly, we were unable to calculate interrater reliability kappa coefficients according to urbanicity, region, facility ownership, and facility size; therefore, interrater reliability sensitivity analyses are not discussed.

Table 6.4 shows percent agreement for Behavioral Signs and Symptoms data elements overall and by setting. Percent agreement was computed for 910 patients/residents: 188 in HHAs, 244 in IRFs, 224 in LTCHs, and 254 in SNFs. Overall percent agreement was high for all Behavioral Signs and Symptoms data elements, ranging from 95 percent to 100 percent. At the setting level, percent agreement was also high for all data elements and ranged from 96 percent to 100 percent in HHAs, 93 percent to 100 percent in IRFs, 97 percent to 100 percent in LTCHs, and 95 percent to 100 percent in SNFs.

**Table 6.4. Interrater Reliability Percent Agreement for Behavioral Signs and Symptoms Data Elements**

| Data Element | HHA (*n* = 188) | IRF (*n* = 244) | LTCH (*n* = 224) | SNF (*n* = 254) | Overall (*n* = 910) |
|---|---|---|---|---|---|
| Physical behavioral symptoms directed toward others (h1a) | 100 | 100 | 100 | 100 | 100 |
| Verbal behavioral symptoms directed toward others (h1b) | 99 | 99 | 99 | 99 | 99 |
| Other behavioral symptoms directed toward others (h1c) | 100 | 100 | 100 | 100 | 100 |
| Did patient/resident reject evaluation or care (h4) | 96 | 93 | 97 | 95 | 95 |

### *Day 3, 5, 7 Repeat Assessment Evaluation*

Table 6.5 summarizes patterns of change for the Behavioral Signs and Symptoms data elements across assessment days. Patterns are characterized as "no change" (scores stay the same across assessment days), "improve" (behavioral symptom goes away across assessment days), "worsen" (behavioral symptom emerges across assessment days), and "fluctuate" (behavioral symptoms come and go across assessment days). As described in more detail in Volume 3, repeat assessment data were collected on a subset of the National Beta Test admission sample of patients/residents according to setting-level target totals. For example, each participating HHA was asked to contribute five patients for repeat assessment. Inclusion in repeat assessment data collection depended on assessor ability to complete the initial assessment on Day 3, availability of the assessor to return for repeat assessments on Days 5 and 7, and willingness of the

patient/resident to complete multiple assessments. For the Behavioral Signs and Symptoms data elements, 473 patients/residents participated in repeat assessments at Days 3, 5, and 7: 93 in HHAs, 122 in IRFs, 71 in LTCHs, and 187 in SNFs. There were no significant overall differences across days for any data elements. Responses showed very little change according to which day data were collected. Overall, there was no change for 99 percent of patients/residents for physical behavioral symptoms, 96 percent for verbal behavioral symptoms, 98 percent for other behavioral symptoms, and 90 percent for rejection of evaluation or care. These general trends were statistically supported.

**Table 6.5. Day 3, 5, and 7 Repeat Assessment Results for Behavioral Signs and Symptoms Data Elements (counts)**

| Data Element | HHA (*n* = 93) | IRF (*n* = 122) | LTCH (*n* = 71) | SNF (*n* = 187) | Overall (*n* = 473) |
|---|---|---|---|---|---|
| Physical behavioral symptoms directed toward others (h1a) | | | | | |
| No change | 93 | 121 | 68 | 186 | 468 |
| Worsen | 0 | 0 | 1 | 0 | 1 |
| Improve | 0 | 0 | 0 | 1 | 1 |
| Fluctuate | 0 | 0 | 0 | 0 | 0 |
| Verbal behavioral symptoms directed toward others (h1b) | | | | | |
| No change | 91 | 117 | 67 | 181 | 456 |
| Worsen | 1 | 1 | 2 | 3 | 7 |
| Improve | 1 | 1 | 0 | 1 | 3 |
| Fluctuate | 0 | 1 | 0 | 1 | 2 |
| Other behavioral symptoms directed toward others (h1c) | | | | | |
| No change | 92 | 119 | 69 | 185 | 465 |
| Worsen | 1 | 0 | 0 | 0 | 1 |
| Improve | 0 | 0 | 0 | 1 | 1 |
| Fluctuate | 0 | 1 | 0 | 0 | 1 |
| Did patient/resident reject evaluation or care (h4) | | | | | |
| No change | 89 | 109 | 59 | 170 | 427 |
| Worsen | 1 | 5 | 3 | 3 | 12 |
| Improve | 0 | 4 | 3 | 6 | 13 |
| Fluctuate | 3 | 4 | 5 | 6 | 18 |

### *Admission to Discharge*

Table 6.6 summarizes patterns of change from admission to discharge for the Behavioral Signs and Symptoms data elements. Patterns are characterized as “no change” (scores stay the

same at admission and discharge), "improve" (behavioral symptom goes away from admission to discharge), and "worsen" (behavioral symptom emerges from admission to discharge). As described in more detail in Volume 3, discharge data were collected on a subset of the National Beta Test admission sample of patients/residents. Availability of discharge data depended on advance notification of discharge and the ability to schedule assessments among the facility staff assessors at each participating site. For Behavioral Signs and Symptoms data elements, admission and discharge data were collected on 770 patients/residents: 138 in HHAs, 333 in IRFs, 82 in LTCHs, and 217 in SNFs. Responses from admission to discharge were consistent for the majority of data elements, with an overall range of 92 percent to 99 percent experiencing no change. Similar to the Day 3, 5, and 7 results, there were no overall statistical differences in responses between admission and discharge.

**Table 6.6. Admission to Discharge Results for Behavioral Signs and Symptoms Data Elements (counts)**

| Data Element | HHA (*n* = 138) | IRF (*n* = 333) | LTCH (*n* = 82) | SNF (*n* = 217) | Overall (*n* = 770) |
|---|---|---|---|---|---|
| Physical behavioral symptoms directed toward others (h1a) | | | | | |
| No change | 135 | 328 | 81 | 216 | 760 |
| Worsen | 1 | 1 | 0 | 0 | 2 |
| Improve | 0 | 2 | 1 | 0 | 3 |
| Verbal behavioral symptoms directed toward others (h1b) | | | | | |
| No change | 132 | 324 | 79 | 212 | 747 |
| Worsen | 1 | 4 | 1 | 1 | 7 |
| Improve | 2 | 4 | 2 | 3 | 11 |
| Other behavioral symptoms directed toward others (h1c) | | | | | |
| No change | 137 | 329 | 81 | 213 | 760 |
| Worsen | 0 | 0 | 1 | 0 | 1 |
| Improve | 0 | 2 | 0 | 2 | 4 |
| Did patient/resident reject evaluation or care (h4) | | | | | |
| No change | 133 | 290 | 76 | 207 | 706 |
| Worsen | 2 | 29 | 3 | 4 | 38 |
| Improve | 2 | 12 | 3 | 6 | 23 |

### *Assessor Feedback*

Assessors mentioned that information about behaviors is very clinically relevant. Facility/agency staff reported that this information is routinely shared in the daily huddle or morning meeting and is documented. In focus groups, Behavioral Signs and Symptoms emerged

as an important safety issue for the staff. Addressing these behaviors is resource-intensive, and the staff's response will depend on whether the behavior represents a health status change:

> We're trying to figure out what's wrong. Is [adverse behavior] because of the illness or what's going on right now? We don't want to throw a pill at it . . . . And then on the tail end you find out, they've been going to the mental health doctor a whole lot of years, and they [were] on this medicine, and now we're not giving it to them, because it's not on the med list. So then, all of a sudden, they're off this medicine, and then their behaviors are out of control.
>
> —Durham, N.C., SNF staff

However, some assessors pointed out that documentation of behavioral signs and symptoms is likely inconsistent. In some cases, PAC staff may hesitate to record behavioral signs and symptoms because having a formal record of disruptive behaviors may prevent a transfer to another PAC provider. Facility/agency staff mentioned that it can be particularly hard to know the frequency of behaviors, even if the behavioral signs and symptoms were documented in the chart. In the assessor survey, facility/agency staff and research nurses scored the Behavioral Signs and Symptoms data elements in the middle to high range of clinical utility, relative to other data elements.

That said, facility/agency staff pointed out that the Behavioral Signs and Symptoms data elements are valuable for effective transfers across PAC settings and that standardization would increase continuity of care and better prepare the staff at the next site of care:

> I feel like a lot of facilities that they're transferring from, it's very unfortunate but a lot of that information is hidden because they want them to be accepted into your facility. So, we may not know they've had these behaviors . . . until they start having all these behaviors within a few days, and then we're really starting to dig for all this information and say, oh, well, they were restrained two days before they came to us. It would have been nice to know that.
>
> —Phoenix, Ariz., SNF staff

## Summary

Results for the Behavioral Signs and Symptoms data elements indicate reasonable overall support for cross-setting standardization. As expected, prevalence rates for all data elements were low, but assessors noted that behavioral problems are an important safety issue for staff, addressing these behaviors is resource-intensive, and the staff's response will depend on whether the behavior represents a health status change. Assessors also noted that documentation of problematic behavior currently might be inconsistent, though documentation of these behaviors would likely improve if the Behavioral Signs and Symptoms data elements are included for cross-setting standardization. Results of the Day 3, 5, and 7 repeat assessments indicate that the data elements are fairly stable regardless of the day of assessment; that is, it did not matter on which day the assessment was administered. Behavior presence was also fairly stable from admission to discharge. We expected that patients exhibiting Behavioral Symptoms would have

longer lengths of stay and require more assistance on ADLs. Although the associations between Behavioral Symptoms and these patient/resident characteristics did not align with our expected results, this was likely because of the very low rate of any patients/residents displaying behavioral symptoms. Results from the National Beta Test support feasibility and interrater reliability, relatively low burden, and high clinical utility of the Behavioral Signs and Symptoms.

# 7. Conclusion

The National Beta Test evaluated several candidate standardized data elements in the clinical category of *cognitive function* for use in the PAC assessment instruments. These data elements included the BIMS, CAM, Expression and Understanding, and Behavioral Signs and Symptoms. It should be noted that rates of cognitive impairment, delirium, inability to express and understand, and problematic behaviors reported in this volume are slightly lower than would be observed in the general PAC patient/resident population because of the National Beta Test design's requirement that individuals be able to communicate meaningfully to be administered these data elements. Noncommunicative patients/residents in participating National Beta Test facilities/agencies were administered a separate assessment protocol designed specifically for noncommunicative patients, which is described in Volume 8.

The general performance of these four data elements is summarized for the combined sample in Table 7.1. As shown in Table 7.1, all four cognitive function data elements performed reasonably well, showing feasibility, acceptable reliability, and substantial support from assessors. However, there are some differences in performance among the four that are worthy of consideration. Specifically, in terms of feasibility, missing data were very low for all four tested data elements, but there was some variability in time to complete. Of the four tested, the BIMS, which consists of seven questions, took the longest to complete (mean = 2.2 minutes, SD = 1.2). Both the CAM and Behavioral Signs and Symptoms took an average of 1.4 minutes to complete (SDs = 0.7 and 0.8, respectively). The three–data element version of Expression and Understanding took 0.8 minutes (SD = 0.4), while the two–data element version took 0.7 minutes (SD = 0.3) overall.

Interrater reliability was acceptable for all data elements, although some performed better than others. Overall kappas for the BIMS were excellent (ranging from 0.83 to 0.93). Kappas were good to moderate for Expression and Understanding, with the three–data element version of Expression and Understanding performing noticeably better than the two–data element version. Kappas, where calculated, were also good to moderate for the CAM. However, several kappas could not be calculated for the CAM, and no kappas were calculable for Behavioral Signs and Symptoms, because of low prevalence rates. All data elements showed acceptable interrater reliability based on percent agreement. Repeat assessment results (not shown in Table 7.1) generally showed very little variability in data element scores. This consistency was true for repeated assessments conducted on the same patient/resident on Days 3, 5, and 7 and at discharge, reinforcing the overall stability of the data elements. Still, the few patients/residents who did show change between admission and discharge imply that assessment of cognitive impairment, delirium, expression and understanding, and problematic behaviors may be

necessary at both admission and discharge to obtain a complete picture of a patient's/resident's cognitive function during his or her PAC stay.

**Table 7.1. Summary of Cognitive Function Data Element Performance in National Beta Test (Combined Sample)**

| Data Element | Time to Complete (Mean, SD) | Interrater Reliability (Kappa) | Interrater Reliability (Percent Agreement) | Assessor Feedback |
|---|---|---|---|---|
| BIMS | 2.2 minutes (1.2 minutes) | 0.83–0.93 | 94–98% | High clinical utility, low burden |
| CAM | 1.4 minutes (0.7 minutes) | 0.66 | 91–96% | Moderate clinical utility, moderately low burden |
| Expression and Understanding (three–data element) | 0.8 minutes (0.4 minutes) | 0.59–0.64 | 93–95% | High clinical utility, very low burden |
| Expression and Understanding (two–data element) | 0.7 minutes (0.3 minutes) | 0.42, 0.32 | 89%, 86% | High clinical utility, very low burden |
| Behavioral Signs and Symptoms | 1.4 minutes (0.8 minutes) | — | 95–100% | Clinically useful, especially for transfers |

NOTES: Kappa not shown for items with proportions out of range for stable kappa estimate (per study power calculations). Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

As with the quantitative results, assessor feedback was generally supportive of all the cognitive function data elements. That is, the data elements were all deemed at least moderately clinically useful and to have reasonably low burden by the clinical assessors in this study. The assessors did raise minor concerns about the frequent assessment of the data elements, which could affect the assessment of short-term memory for the recall of three words in that the same words are always used, and patients/residents may be using their long-term memory skills from a previous assessment instead. Assessors also noted a tendency for Behavioral Signs and Symptoms documentation to be somewhat inconsistent, mostly because presence of these symptoms could influence a patient's/resident's opportunity to transfer to another setting. However, assessors did not consider these issues critical enough to outweigh their overall support for these data elements.

# Appendix. Supplementary Tables

## Supplementary Tables for BIMS

**Table A.1. Frequencies for BIMS Impairment Categorization by Patient/Resident Characteristics and Clinical Groups in the HHA Setting (percent)**

| Patient/Resident Characteristics and Clinical Groups | Intact | Moderately Impaired | Severely Impaired |
|---|---|---|---|
| Gender (*n* = 602) | | | |
| Male (*n* = 211) | 77 | 19 | 4 |
| Female (*n* = 391) | 81 | 16 | 3 |
| Age (*n* = 598)[a] | | | |
| 18–44 (*n* = 2) | 50 | 50 | 0 |
| 45–64 (*n* = 59) | 90 | 10 | 0 |
| 65–74 (*n* = 172) | 90 | 9 | 1 |
| 75–89 (*n* = 299) | 75 | 20 | 5 |
| 90 or older (*n* = 66) | 62 | 30 | 8 |
| Length of stay (*n* = 481; mean, SD) | 30.7 (15.3) | 30.6 (17.3) | 34.8 (15.2) |
| Disposition at discharge (*n* = 594) | | | |
| Home (*n* = 442) | 81 | 16 | 3 |
| Hospital (*n* = 20) | 80 | 20 | 0 |
| Hospice (*n* = 3810 | 90 | 0 | 10 |
| HHA (*n* = 15) | 80 | 20 | 0 |
| IRF (*n* = 4) | 25 | 50 | 25 |
| LTCH (*n* = 1) | 100 | 0 | 0 |
| SNF (*n* = 6) | 83 | 17 | 0 |
| Other (*n* = 96) | 72 | 23 | 5 |
| Clinical conditions (*n* = 405) | | | |
| Sepsis | | | |
| Yes (*n* = 9) | 78 | 11 | 11 |
| No (*n* = 396) | 80 | 17 | 3 |
| Heart failure | | | |
| Yes (*n* = 32) | 91 | 9 | 0 |
| No (*n* = 373) | 79 | 18 | 3 |
| Stroke | | | |
| Yes (*n* = 7) | 71 | 29 | 0 |
| No (*n* = 398) | 80 | 17 | 3 |

| Patient/Resident Characteristics and Clinical Groups | Intact | Moderately Impaired | Severely Impaired |
|---|---|---|---|
| Mobility—Transfer from lying to sitting (*n* = 380) | | | |
| Independent (*n* = 28) | 86 | 11 | 4 |
| Setup or clean-up assistance (*n* = 59) | 75 | 22 | 3 |
| Supervision or touching assistance (*n* = 117) | 83 | 16 | 1 |
| Partial/moderate assistance (*n* = 119) | 82 | 16 | 2 |
| Substantial/maximal assistance (*n* = 51) | 73 | 20 | 8 |
| Dependent (*n* = 6) | 83 | 0 | 17 |

[a] Significant ($p < 0.05$) associations with BIMS impairment category as indicated by chi-square tests of independence (ANOVA for length of stay).

**Table A.2. Frequencies for BIMS Impairment Categorization by Patient/Resident Characteristics and Clinical Groups in the IRF Setting (percent)**

| Patient/Resident Characteristics and Clinical Groups | Intact | Moderately Impaired | Severely Impaired |
|---|---|---|---|
| Gender (*n* = 735)[a] | | | |
| Male (*n* = 315) | 77 | 18 | 5 |
| Female (*n* = 420) | 86 | 12 | 1 |
| Age (*n* = 732)[a] | | | |
| 18–44 (*n* = 6) | 100 | 0 | 0 |
| 45–64 (*n* = 57) | 84 | 14 | 2 |
| 65–74 (*n* = 283) | 88 | 9 | 3 |
| 75–89 (*n* = 331) | 78 | 20 | 2 |
| 90 or older (*n* = 55) | 78 | 16 | 5 |
| Length of stay (*n* = 716; mean, SD)[a] | 13.9 (5.0) | 14.7 (5.4) | 16.5 (4.7) |
| Disposition at discharge (*n* = 730)[a] | | | |
| Home (*n* = 315) | 86 | 10 | 4 |
| Hospital (*n* = 37) | 73 | 24 | 3 |
| Hospice (*n* = 7) | 71 | 14 | 14 |
| HHA (*n* = 247) | 83 | 16 | 2 |
| IRF (*n* = 1) | 100 | 0 | 0 |
| LTCH (*n* = 1) | 0 | 0 | 100 |
| SNF (*n* = 102) | 76 | 23 | 1 |
| Other (*n* = 20) | 85 | 15 | 0 |
| Clinical conditions (*n* = 574) | | | |
| Sepsis | | | |
| Yes (*n* = 23) | 87 | 13 | 0 |
| No (*n* = 551) | 82 | 15 | 3 |

| Patient/Resident Characteristics and Clinical Groups | Intact | Moderately Impaired | Severely Impaired |
|---|---|---|---|
| Heart failure | | | |
| Yes (*n* = 128) | 80 | 17 | 2 |
| No (*n* = 446) | 83 | 14 | 2 |
| Stroke | | | |
| Yes (*n* = 99)[a] | 74 | 22 | 4 |
| No (*n* = 475) | 84 | 13 | 2 |
| Hygiene—Toileting (*n* = 558) | | | |
| Independent (*n* = 5) | 100 | 0 | 0 |
| Setup or clean-up assistance (*n* = 22) | 86 | 14 | 0 |
| Supervision or touching assistance (*n* = 121) | 83 | 16 | 2 |
| Partial/moderate assistance (*n* = 133) | 87 | 11 | 2 |
| Substantial/maximal assistance (*n* = 136) | 81 | 15 | 4 |
| Dependent (*n* = 141) | 79 | 19 | 1 |
| Mobility—transfer from lying to sitting (*n* = 571) | | | |
| Independent (*n* = 42) | 93 | 5 | 2 |
| Setup or clean-up assistance (*n* = 16) | 94 | 6 | 0 |
| Supervision or touching assistance (*n* = 184) | 83 | 16 | 1 |
| Partial/moderate assistance (*n* = 214) | 82 | 15 | 3 |
| Substantial/maximal assistance (*n* = 90) | 81 | 14 | 4 |
| Dependent (*n* = 25) | 76 | 24 | 0 |

[a] Significant ($p < 0.05$) associations with BIMS impairment category as indicated by chi-square tests of independence (ANOVA for length of stay).

**Table A.3. Frequencies for BIMS Impairment Categorization by Patient/Resident Characteristics and Clinical Groups in the LTCH Setting (percent)**

| Patient/Resident Characteristics and Clinical Groups | Intact | Moderately Impaired | Severely Impaired |
|---|---|---|---|
| Gender (*n* = 454) | | | |
| Male (*n* = 232) | 72 | 21 | 7 |
| Female (*n* = 222) | 75 | 17 | 8 |
| Age (*n* = 455)[a] | | | |
| 18–44 (*n* = 21) | 95 | 5 | 0 |
| 45–64 (*n* = 115) | 83 | 14 | 3 |
| 65–74 (*n* = 160) | 76 | 19 | 5 |
| 75–89 (*n* = 145) | 63 | 24 | 12 |
| 90 or older (*n* = 14) | 50 | 29 | 21 |
| Length of stay (*n* = 400; mean, SD) | 23.6 (11.0) | 24.2 (11.5) | 23.0 (11.0) |

| Patient/Resident Characteristics and Clinical Groups | Intact | Moderately Impaired | Severely Impaired |
|---|---|---|---|
| Disposition at discharge (*n* = 440)[a] | | | |
| Home (*n* = 89) | 83 | 11 | 6 |
| Hospital (*n* = 31) | 84 | 10 | 6 |
| Hospice (*n* = 12) | 50 | 50 | 0 |
| HHA (*n* = 76) | 80 | 17 | 3 |
| IRF (*n* = 43) | 77 | 14 | 9 |
| LTCH (*n* = 1) | 0 | 100 | 0 |
| SNF (*n* = 122) | 67 | 20 | 12 |
| Other (*n* = 66) | 64 | 30 | 6 |
| Clinical conditions (*n* = 389) | | | |
| Sepsis | | | |
| Yes (*n* = 66) | 71 | 18 | 11 |
| No (*n* = 323) | 75 | 18 | 7 |
| Heart Failure | | | |
| Yes (*n* = 13) | 92 | 8 | 0 |
| No (*n* = 376) | 74 | 19 | 7 |
| Stroke | | | |
| Yes (*n* = 29)[a] | 52 | 14 | 34 |
| No (*n* = 360) | 76 | 19 | 5 |
| Hygiene—Toileting (*n* = 377)[a] | | | |
| Independent (*n* = 42) | 90 | 10 | 0 |
| Setup or clean-up assistance (*n* = 33) | 88 | 12 | 0 |
| Supervision or touching assistance (*n* = 57) | 82 | 12 | 5 |
| Partial/moderate assistance (*n* = 50) | 80 | 16 | 4 |
| Substantial/maximal assistance (*n* = 62) | 73 | 19 | 8 |
| Dependent (*n* = 133) | 63 | 24 | 13 |
| Mobility—Transfer from lying to sitting (*n* = 333)[a] | | | |
| Independent (*n* = 59) | 90 | 10 | 0 |
| Setup or clean-up assistance (*n* = 23) | 91 | 9 | 0 |
| Supervision or touching assistance (*n* = 58) | 81 | 16 | 3 |
| Partial/moderate assistance (*n* = 69) | 81 | 16 | 3 |
| Substantial/maximal assistance (*n* = 50) | 66 | 22 | 12 |
| Dependent (*n* = 74) | 62 | 23 | 15 |

[a] Significant ($p < 0.05$) associations with BIMS impairment category as indicated by chi-square tests of independence (ANOVA for length of stay).

**Table A.4. Frequencies for BIMS Impairment Categorization by Patient/Resident Characteristics and Clinical Groups in the SNF Setting (percent)**

| Patient/Resident Characteristics and Clinical Groups | Intact | Moderately Impaired | Severely Impaired |
|---|---|---|---|
| Gender (*n* = 1,075) | | | |
| Male (*n* = 417) | 69 | 25 | 7 |
| Female (*n* = 658) | 73 | 20 | 7 |
| Age (*n* = 1,071)[a] | | | |
| 18–44 (*n* = 9) | 89 | 11 | 0 |
| 45–64 (*n* = 75) | 84 | 16 | 0 |
| 65–74 (*n* = 281) | 77 | 21 | 2 |
| 75–89 (*n* = 536) | 69 | 23 | 9 |
| 90 or older (*n* = 170) | 65 | 24 | 11 |
| Length of stay (*n* = 932; mean, SD)[a] | 21.0 (12.3) | 22.0 (12.0) | 26.6 (13.7) |
| Disposition at discharge (*n* = 1,054)[a] | | | |
| Home (*n* = 465) | 75 | 21 | 4 |
| Hospital (*n* = 109) | 62 | 27 | 11 |
| Hospice (*n* = 9) | 78 | 22 | 0 |
| SNF (*n* = 43) | 65 | 26 | 9 |
| IRF (*n* = 1) | 100 | 0 | 0 |
| HHA (*n* = 274) | 77 | 19 | 4 |
| LTCH (*n* = 10) | 40 | 50 | 10 |
| Other (*n* = 143) | 64 | 19 | 17 |
| Clinical conditions (*n* = 847) | | | |
| Sepsis | | | |
| Yes (*n* = 52) | 67 | 23 | 10 |
| No (*n* = 795) | 72 | 22 | 7 |
| Heart failure | | | |
| Yes (*n* = 209) | 66 | 27 | 8 |
| No (*n* = 638) | 73 | 20 | 7 |
| Stroke | | | |
| Yes (*n* = 65)[a] | 54 | 32 | 14 |
| No (*n* = 782) | 73 | 21 | 6 |

| Patient/Resident Characteristics and Clinical Groups | Intact | Moderately Impaired | Severely Impaired |
|---|---|---|---|
| Hygiene—Toileting (*n* = 570)[a] | | | |
| Independent (*n* = 22) | 77 | 23 | 0 |
| Setup or clean-up assistance (*n* = 23) | 87 | 13 | 0 |
| Supervision or touching assistance (*n* = 139) | 78 | 17 | 5 |
| Partial/moderate assistance (*n* = 177) | 77 | 17 | 6 |
| Substantial/maximal assistance (*n* = 132) | 64 | 30 | 7 |
| Dependent (*n* = 77) | 65 | 22 | 13 |
| Mobility—Transfer from lying to sitting (*n* = 562) | | | |
| Independent (*n* = 57) | 74 | 25 | 2 |
| Setup or clean-up assistance (*n* = 14) | 79 | 14 | 7 |
| Supervision or touching assistance (*n* = 164) | 75 | 20 | 5 |
| Partial/moderate assistance (*n* = 199) | 73 | 20 | 7 |
| Substantial/maximal assistance (*n* = 101) | 72 | 22 | 6 |
| Dependent (*n* = 27) | 52 | 26 | 22 |

[a] Significant ($p < 0.05$) associations with BIMS impairment category as indicated by chi-square tests of independence (ANOVA for length of stay).

**Table A.5. Time to Complete the BIMS by Urbanicity (minutes)**

| Characteristic | Urban (*n* = 1,695) | Nonurban (*n* = 113) | Overall (*n* = 1,808) |
|---|---|---|---|
| Mean (SD) | 2.2 (1.2) | 2.3 (1.0) | 2.2 (1.2) |

**Table A.6. Time to Complete the BIMS by Region (minutes)**

| Characteristic | Northeast (*n* = 486) | South (*n* = 668) | Midwest (*n* = 367) | West (*n* = 287) | Overall (*n* = 1,808) |
|---|---|---|---|---|---|
| Mean (SD) | 2.1 (1.1) | 2.2 (1.2) | 2.3 (1.3) | 2.1 (1.1) | 2.2 (1.2) |

**Table A.7. Time to Complete the BIMS by Facility Ownership (minutes)**

| Characteristic | For-Profit (*n* = 678) | Nonprofit (*n* = 1,114) | Overall (*n* = 1,808) |
|---|---|---|---|
| Mean (SD) | 2.1 (1.1) | 2.2 (1.2) | 2.2 (1.2) |

NOTE: Patient/resident numbers in for-profit and nonprofit categories do not sum to overall total because of missing profit status data.

### Table A.8. Time to Complete the BIMS by Facility Size (minutes)

| Characteristic | Below Setting-Type Median (*n* = 757) | Above Setting-Type Median (*n* = 1,050) | Overall (*n* = 1,808) |
|---|---|---|---|
| Mean (SD) | 2.1 (1.1) | 2.2 (1.2) | 2.2 (1.2) |

NOTE: Patient/resident numbers in below and above setting-type median categories do not sum to overall total because of missing facility-size data.

### Table A.9. Interrater Reliability Kappa or Weighted Kappa for BIMS Data Elements by Urbanicity

| Data Element | Urban (*n* = 899) | Nonurban (*n* = 67) |
|---|---|---|
| Number of words repeated after first attempt (b1a) | 0.71 | 1.00 |
| Patient able to recall current year (b1b) | 0.90 | 1.00 |
| Patient able to recall current month (b1c) | 0.90 | 0.72 |
| Patient able to recall current day of week (b1d) | 0.87 | 0.94 |
| Recalls “sock” (b1e) | 0.90 | 0.95 |
| Recalls “blue” (b1f) | 0.83 | 0.79 |
| Recalls “bed” (b1g) | 0.93 | 0.92 |
| BIMS (categorical) | 0.90 | 1.00 |

### Table A.10. Interrater Reliability Kappa or Weighted Kappa for BIMS Data Elements by Region

| Data Element | Northeast (*n* = 219) | South (*n* = 365) | Midwest (*n* = 210) | West (*n* = 182) |
|---|---|---|---|---|
| Number of words repeated after first attempt (b1a) | 0.72 | 0.61 | 0.82 | 0.79 |
| Patient able to recall current year (b1b) | 0.88 | 0.92 | 0.96 | 0.81 |
| Patient able to recall current month (b1c) | 0.91 | 0.86 | 0.98 | 0.84 |
| Patient able to recall current day of week (b1d) | 0.90 | 0.93 | 0.82 | 0.79 |
| Recalls “sock” (b1e) | 0.89 | 0.91 | 0.89 | 0.93 |
| Recalls “blue” (b1f) | 0.83 | 0.81 | 0.87 | 0.85 |
| Recalls “bed” (b1g) | 0.93 | 0.93 | 0.91 | 0.90 |
| BIMS (categorical) | 0.90 | 0.92 | 0.93 | 0.83 |

NOTE: Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

**Table A.11. Interrater Reliability Kappa or Weighted Kappa for BIMS Data Elements by Facility Ownership**

| Data Element | For-Profit (*n* = 611) | Nonprofit (*n* = 349) |
|---|---|---|
| Number of words repeated after first attempt (b1a) | — | — |
| Recalls current year (b1b) | 0.91 | 0.89 |
| Recalls current month (b1c) | 0.92 | — |
| Recalls current day of week (b1d) | 0.90 | 0.84 |
| Recalls “sock” (b1e) | 0.91 | 0.90 |
| Recalls “blue” (b1f) | 0.87 | 0.76 |
| Recalls “bed” (b1g) | 0.94 | 0.90 |
| BIMS impairment category (based on responses to b1a–b1g) | 0.92 | 0.87 |

NOTE: Interrater reliability not shown for data elements with proportions out of range for stable kappa estimate (per study power calculations). Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

**Table A.12. Interrater Reliability Kappa or Weighted Kappa for BIMS Data Elements by Facility Size**

| Data Element | Below Setting-Type Median (*n* = 426) | Above Setting-Type Median (*n* = 539) |
|---|---|---|
| Number of words repeated after first attempt (b1a) | — | — |
| Recalls current year (b1b) | 0.91 | 0.90 |
| Recalls current month (b1c) | — | 0.92 |
| Recalls current day of week (b1d) | 0.89 | 0.87 |
| Recalls “sock” (b1e) | 0.92 | 0.89 |
| Recalls “blue” (b1f) | 0.86 | 0.81 |
| Recalls “bed” (b1g) | 0.94 | 0.92 |
| BIMS impairment category (based on responses to b1a–b1g) | 0.93 | 0.89 |

NOTE: Interrater reliability not shown for data elements with proportions out of range for stable kappa estimate (per study power calculations). Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

## Supplementary Tables for CAM

**Table A.13. Frequencies for the CAM Change in Mental Status Data Element by Patient/Resident Characteristics and Clinical Groups in the HHA Setting (percent)**

| Patient/Resident Characteristics and Clinical Groups | Change in Mental Status (Yes) |
|---|---|
| Gender (*n* = 608) | |
| Male (*n* = 213) | 7 |
| Female (*n* = 395) | 4 |
| Age (*n* = 605) | |
| 18–44 (*n* = 3) | 0 |
| 45–64 (*n* = 59) | 7 |
| 65–74 (*n* = 172) | 4 |
| 75–89 (*n* = 302) | 5 |
| 90 or older (*n* = 69) | 4 |
| Length of stay (*n* = 485; mean, SD) | Yes: 32.0 (15.9)<br>No: 31.2 (15.6) |
| Disposition at discharge (*n* = 600) | |
| Home (*n* = 443) | 4 |
| Hospital (*n* = 23) | 4 |
| Hospice (*n* = 12) | 17 |
| HHA (*n* = 15) | 13 |
| IRF (*n* = 4) | 0 |
| LTCH (*n* = 1) | 0 |
| SNF (*n* = 6) | 0 |
| Other (*n* = 96) | 6 |
| Clinical conditions (*n* = 408) | |
| Sepsis | |
| Yes (*n* = 7) | 0 |
| No (*n* = 401) | 5 |
| Heart failure | |
| Yes (*n* = 32) | 3 |
| No (*n* = 376) | 5 |
| Stroke | |
| Yes (*n* = 6) | 17 |
| No (*n* = 402) | 5 |

| Patient/Resident Characteristics and Clinical Groups | Change in Mental Status (Yes) |
|---|---|
| Mobility—Transfer from lying to sitting (*n* = 382) | |
| Independent (*n* = 29) | 3 |
| Setup or clean-up assistance (*n* = 57) | 7 |
| Supervision or touching assistance (*n* = 116) | 4 |
| Partial/moderate assistance (*n* = 124) | 5 |
| Substantial/maximal assistance (*n* = 50) | 4 |
| Dependent (*n* = 6) | 17 |

**Table A.14. Frequencies for the CAM Change in Mental Status Data Element by Patient/Resident Characteristics and Clinical Groups in the IRF Setting (percent)**

| Patient/Resident Characteristics and Clinical Groups | Change in Mental Status (Yes) |
|---|---|
| Gender (*n* = 735) | |
| Male (*n* = 319) | 8 |
| Female (*n* = 416) | 5 |
| Age (*n* = 732) | |
| 18–44 (*n* = 5) | 0 |
| 45–64 (*n* = 59) | 7 |
| 65–74 (*n* = 285) | 7 |
| 75–89 (*n* = 330) | 7 |
| 90 or older (*n* = 53) | 0 |
| Length of stay (*n* = 717; mean, SD)[a] | Yes: 16.0 (5.2)<br>No: 14.0 (5.0) |
| Disposition at discharge (*n* = 731) | |
| Home (*n* = 314) | 8 |
| Hospital (*n* = 37) | 8 |
| Hospice (*n* = 7) | 29 |
| HHA (*n* = 250) | 4 |
| IRF (*n* = 1) | 0 |
| LTCH (*n* = 0) | N/A |
| SNF (*n* = 102) | 10 |
| Other (*n* = 20) | 0 |
| Clinical conditions (*n* = 569) | |
| Sepsis | |
| Yes (*n* = 24) | 4 |
| No (*n* = 545) | 8 |

| Patient/Resident Characteristics and Clinical Groups | Change in Mental Status (Yes) |
|---|---|
| Heart failure | |
| Yes (*n* = 127) | 8 |
| No (*n* = 442) | 7 |
| Stroke | |
| Yes (*n* = 97)[a] | 13 |
| No (*n* = 472) | 6 |
| Hygiene—Toileting (*n* = 553) | |
| Independent (*n* = 5) | 0 |
| Setup or clean-up assistance (*n* = 20) | 10 |
| Supervision or touching assistance (*n* = 117) | 6 |
| Partial/moderate assistance (*n* = 135) | 5 |
| Substantial/maximal assistance (*n* = 139) | 9 |
| Dependent (*n* = 137) | 10 |
| Mobility—Transfer from lying to sitting (*n* = 566) | |
| Independent (*n* = 42) | 2 |
| Setup or clean-up assistance (*n* = 16) | 13 |
| Supervision or touching assistance (*n* = 182) | 5 |
| Partial/moderate assistance (*n* = 212) | 8 |
| Substantial/maximal assistance (*n* = 88) | 9 |
| Dependent (*n* = 26) | 15 |

NOTE: N/A = not applicable.

**Table A.15. Frequencies for the CAM Change in Mental Status Data Element by Patient/Resident Characteristics and Clinical Groups in the LTCH Setting (percent)**

| Patient/Resident Characteristics and Clinical Groups | Change in Mental Status (Yes) |
|---|---|
| Gender (*n* = 450) | |
| Male (*n* = 231) | 5 |
| Female (*n* = 219) | 4 |
| Age (*n* = 451) | |
| 18–44 (*n* = 20) | 0 |
| 45–64 (*n* = 114) | 4 |
| 65–74 (*n* = 162) | 4 |
| 75–89 (*n* = 141) | 6 |
| 90 or older (*n* = 14) | 7 |
| Length of stay (*n* = 399; mean, SD) | Yes: 23.4 (11.3)<br>No: 24.0 (11.1) |

| Patient/Resident Characteristics and Clinical Groups | Change in Mental Status (Yes) |
|---|---|
| Disposition at discharge (*n* = 436) | |
| Home (*n* = 88) | 5 |
| Hospital (*n* = 29) | 0 |
| Hospice (*n* = 12) | 8 |
| HHA (*n* = 76) | 1 |
| IRF (*n* = 45) | 7 |
| LTCH (*n* = 1) | 0 |
| SNF (*n* = 123) | 4 |
| Other (*n* = 62) | 10 |
| Clinical conditions (*n* = 383) | |
| Sepsis | |
| Yes (*n* = 66) | 9 |
| No (*n* = 317) | 4 |
| Heart failure | |
| Yes (*n* = 13) | 0 |
| No (*n* = 370) | 5 |
| Stroke | |
| Yes (*n* = 26) | 8 |
| No (*n* = 357) | 5 |
| Hygiene—Toileting (*n* = 371) | |
| Independent (*n* = 45) | 2 |
| Setup or clean-up assistance (*n* = 33) | 12 |
| Supervision or touching assistance (*n* = 57) | 4 |
| Partial/moderate assistance (*n* = 49) | 4 |
| Substantial/maximal assistance (*n* = 59) | 2 |
| Dependent (*n* = 128) | 7 |
| Mobility—Transfer from lying to sitting (*n* = 333) | |
| Independent (*n* = 62) | 2 |
| Setup or clean-up assistance (*n* = 23) | 13 |
| Supervision or touching assistance (*n* = 58) | 5 |
| Partial/moderate assistance (*n* = 70) | 4 |
| Substantial/maximal assistance (*n* = 49) | 4 |
| Dependent (*n* = 71) | 6 |

**Table A.16. Frequencies for the CAM Change in Mental Status Data Element by Patient/Resident Characteristics and Clinical Groups in the SNF Setting (percent)**

| Patient/Resident Characteristics and Clinical Groups | Change in Mental Status (Yes) |
|---|---|
| Gender (*n* = 1,069) | |
| Male (*n* = 416) | 5 |
| Female (*n* = 653) | 3 |
| Age (*n* = 1,064) | |
| 18–44 (*n* = 9) | 0 |
| 45–64 (*n* = 72) | 4 |
| 65–74 (*n* = 281) | 4 |
| 75–89 (*n* = 533) | 3 |
| 90 or older (*n* = 169) | 5 |
| Length of stay (*n* = 929; mean, SD) | Yes: 19.5 (9.3)<br>No: 21.4 (12.3) |
| Disposition at discharge (*n* = 1,050) | |
| Home (*n* = 470) | 3 |
| Hospital (*n* = 102) | 4 |
| Hospice (*n* = 9) | 0 |
| HHA (*n* = 274) | 1 |
| IRF (*n* = 1) | 0 |
| LTCH (*n* = 10) | 10 |
| SNF (*n* = 43) | 5 |
| Other (*n* = 141) | 7 |
| Clinical conditions (*n* = 839) | |
| Sepsis | |
| Yes (*n* = 51) | 4 |
| No (*n* = 788) | 4 |
| Heart failure | |
| Yes (*n* = 201a) | 6 |
| No (*n* = 638) | 3 |
| Stroke | |
| Yes (*n* = 63) | 8 |
| No (*n* = 776) | 3 |

| Patient/Resident Characteristics and Clinical Groups | Change in Mental Status (Yes) |
|---|---|
| Hygiene—Toileting (*n* = 565) | |
| Independent (*n* = 22) | 5 |
| Setup or clean-up assistance (*n* = 23) | 4 |
| Supervision or touching assistance (*n* = 141) | 1 |
| Partial/moderate assistance (*n* = 173) | 5 |
| Substantial/maximal assistance (*n* = 132) | 5 |
| Dependent (*n* = 74) | 4 |
| Mobility—Transfer from lying to sitting (*n* = 557) | |
| Independent (*n* = 58) | 7 |
| Setup or clean-up assistance (*n* = 14) | 0 |
| Supervision or touching assistance (*n* = 163) | 2 |
| Partial/moderate assistance (*n* = 202) | 3 |
| Substantial/maximal assistance (*n* = 95) | 3 |
| Dependent (*n* = 25) | 12 |

[a] Significant ($p < 0.05$) associations with change in mental status as indicated by chi-square tests of independence (ANOVA for length of stay).

**Table A.17. Time to Complete the CAM Data Elements by Urbanicity (minutes)**

| Characteristic | Urban (*n* = 1,446) | Nonurban (*n* = 90) | Overall (*n* = 1,536) |
|---|---|---|---|
| Mean (SD) | 1.4 (0.7) | 1.7 (0.7) | 1.4 (0.7) |

**Table A.18. Time to Complete the CAM Data Elements by Region (minutes)**

| Characteristic | Northeast (*n* = 413) | South (*n* = 577) | Midwest (*n* = 301) | West (*n* = 245) | Overall (*n* = 1,536) |
|---|---|---|---|---|---|
| Mean (SD) | 1.4 (0.6) | 1.4 (0.7) | 1.5 (0.7) | 1.4 (0.7) | 1.4 (0.7) |

**Table A.19. Time to Complete the CAM Data Elements by Facility Ownership (minutes)**

| Characteristic | For-Profit (*n* = 550) | Nonprofit (*n* = 979) | Overall (*n* = 1,536) |
|---|---|---|---|
| Mean (SD) | 1.4 (0.6) | 1.4 (0.7) | 1.4 (0.7) |

NOTE: Patient/resident numbers in for-profit and nonprofit categories do not sum to overall total because of missing profit status data.

### Table A.20. Time to Complete the CAM Data Elements by Facility Size

| Characteristic | Below Setting-Type Median (*n* = 652) | Above Setting-Type Median (*n* = 883) | Overall (*n* = 1,536) |
|---|---|---|---|
| Mean (SD) | 1.5 (0.7) | 1.3 (0.6) | 1.4 (0.7) |

NOTE: Patient/resident numbers in below and above setting-type median categories do not sum to overall total because of missing facility-size data.

### Table A.21. Interrater Reliability Kappa or Weighted Kappa for CAM Data Elements by Urbanicity

| Data Element | Urban (*n* = 849) | Nonurban (*n* =65) |
|---|---|---|
| Evidence of change in mental status from baseline (b2a) | 0.66 | 0.82 |
| Did patient have difficulty focusing attention (b2b) | 0.66 | 0.76 |
| Was patient thinking disorganized (b2c) | 0.57 | 0.85 |
| Did patient have altered consciousness (b2d) | 0.66 | 0.59 |

NOTE: Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

### Table A.22. Interrater Reliability Kappa or Weighted Kappa for CAM Data Elements by Region

| Data Element | Northeast (*n* = 198) | South (*n* = 347) | Midwest (*n* = 200) | West (*n* = 169) |
|---|---|---|---|---|
| Evidence of change in mental status from baseline (b2a) | 0.75 | 0.50 | 0.75 | 0.67 |
| Did patient have difficulty focusing attention (b2b) | 0.68 | 0.62 | 0.81 | 0.53 |
| Was patient thinking disorganized (b2c) | 0.59 | 0.64 | 0.71 | 0.19 |
| Did patient have altered consciousness (b2d) | 0.63 | 0.65 | 0.75 | 0.56 |

NOTE: Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

### Table A.23. Interrater Reliability Kappa or Weighted Kappa for CAM Data Elements by Facility Ownership

| Data Element | For-Profit (*n* = 578) | Nonprofit (*n* = 331) |
|---|---|---|
| Evidence of change in mental status from baseline (b2a) | — | — |
| Did patient have difficulty focusing attention (b2b) | 0.68 | 0.63 |
| Was patient thinking disorganized (b2c) | — | — |
| Did patient have altered consciousness (b2d) | — | — |

NOTE: Interrater reliability not shown for data elements with proportions out of range for stable kappa estimate (per study power calculations). Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

**Table A.24. Interrater Reliability Kappa or Weighted Kappa for CAM Data Elements by Facility Size**

| Data Element | Below Setting-Type Median (*n* = 411) | Above Setting-Type Median (*n* = 502) |
|---|---|---|
| Evidence of change in mental status from baseline (b2a) | — | — |
| Did patient have difficulty focusing attention (b2b) | 0.68 | 0.65 |
| Was patient thinking disorganized (b2c) | — | 0.62 |
| Did patient have altered consciousness (b2d) | — | — |

NOTE: Interrater reliability not shown for data elements with proportions out of range for stable kappa estimate (per study power calculations). Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

# Supplementary Tables for Expression and Understanding

**Table A.25. Frequencies for Expression and Understanding Data Elements by Patient/Resident Characteristics and Clinical Groups in the HHA Setting (percent)**

| Patient/Resident Characteristics and Clinical Groups | Expresses Without Difficulty(Yes) | Understands(Yes) |
|---|---|---|
| Gender (*nE* = 620; *nU* = 621) | | |
| Male (*nE* = 223; *nU* = 225) | 84.3 | 81.3 |
| Female (*nE* = 397; *nU* = 396) | 83.4 | 78.0 |
| Age (*nE* = 617[a]; *nU* = 618[a]) | | |
| 18–44 (*nE* = 4; *nU* = 4) | 75.0 | 50.0 |
| 45–64 (*nE* = 59; *nU* = 59) | 78.0 | 76.3 |
| 65–74 (*nE* = 172; *nU* = 173) | 90.7 | 90.8 |
| 75–89 (*nE* = 311; *nU* = 311) | 82.6 | 78.1 |
| 90 or older (*nE* = 71; *nU* = 71) | 76.1 | 59.2 |
| Length of stay (*nE* = 494[a]; *nU* = 485[a]; mean, SD) | Yes: 29.8 (15.2)<br>No: 39.0 (16.2) | Yes: 29.7 (15.1)<br>No: 37.0 (16.8) |
| Disposition at discharge (*nE* = 612[a]; *nU* = 613) | | |
| Home (*nE* = 455; *nU* = 455) | 85.1 | 79.6 |
| Hospital (*nE* = 22; *nU* = 22) | 95.5 | 95.5 |
| Hospice (*nE* = 12; *nU* = 12) | 91.7 | 66.7 |
| HHA (*nE* = 15; *nU* = 15) | 73.3 | 73.3 |
| IRF (*nE* = 4; *nU* = 4) | 50.0 | 50.0 |
| LTCH (*nE* = 1; *nU* = 1) | 0 | 100 |
| SNF (*nE* = 6; *nU* = 6) | 83.3 | 100 |
| Other (*nE* = 97; *nU* = 98) | 77.3 | 75.5 |

| Patient/Resident Characteristics and Clinical Groups | Expresses Without Difficulty(Yes) | Understands(Yes) |
|---|---|---|
| Clinical conditions (*nE* = 416; *nU* = 416) | | |
| Sepsis | | |
| Yes (*nE* = 9; *nU* = 9) | 100 | 88.9 |
| No (*nE* = 407; *nU* = 407) | 82.8 | 79.4 |
| Heart failure | | |
| Yes (*nE* = 32; *nU* = 32) | 90.6 | 81.3 |
| No (*nE* = 384; *nU* = 384) | 82.6 | 79.4 |
| Stroke | | |
| Yes (*nE* = 7; *nU* = 7) | 57.1 | 71.4 |
| No (*nE* = 409; *nU* = 409) | 83.6 | 79.7 |
| Mobility—Transfer from lying to sitting (*nE* = 389[a]; *nU* = 389[a]) | | |
| Independent (*nE* = 30; *nU* = 30) | 86.7 | 80.0 |
| Setup or clean-up assistance (*nE* = 59; *nU* = 59) | 84.8 | 88.1 |
| Supervision or touching assistance (*nE* = 117; *nU* = 117) | 88.9 | 84.6 |
| Partial/moderate assistance (*nE* = 124; *nU* = 124) | 85.5 | 83.1 |
| Substantial/maximal assistance (*nE* = 53; *nU* = 53) | 62.3 | 58.5 |
| Dependent (*nE* = 6; *nU* = 6) | 83.3 | 83.3 |

NOTE: Because of differences in sample sizes for Expression and Understanding, we report sample sizes for each (*nE* = Expression; *nU* = Understanding).
[a] Significant ($p < 0.05$) associations with expression and understanding as indicated by chi-square tests of independence.

**Table A.26. Frequencies for Expression and Understanding Data Elements by Patient/Resident Characteristics and Clinical Groups in the IRF Setting (percent)**

| Patient/Resident Characteristics and Clinical Groups | Expresses Without Difficulty(Yes) | Understands(Yes) |
|---|---|---|
| Gender (*nE* = 748; *nU* = 748) | | |
| Male (*nE* = 321; *nU* = 321) | 89.4 | 89.4 |
| Female (*nE* = 427; *nU* = 427) | 93.2 | 92.5 |
| Age (*nE* = 745; *nU* = 745) | | |
| 18–44 (*nE* = 5; *nU* = 5) | 100 | 100 |
| 45–64 (*nE* = 57; *nU* = 57) | 93.0 | 96.5 |
| 65–74 (*nE* = 290; *nU* = 290) | 93.8 | 90.7 |
| 75–89 (*nE* = 338; *nU* = 338) | 90.2 | 90.8 |
| 90 or older (*nE* = 55; *nU* = 55) | 87.3 | 89.1 |
| Length of stay (*nE* = 728[a]; *nU* = 728[a]; mean, SD) | Yes: 14.0 (5.0)<br>No:15.7 (5.8) | Yes: 14.0 (5.0)<br>No: 15.4 (5.6) |

| Patient/Resident Characteristics and Clinical Groups | Expresses Without Difficulty(Yes) | Understands(Yes) |
|---|---|---|
| Disposition at discharge (*nE* = 742[a]; *nU* = 742) | | |
| Home (*nE* = 317; *nU* = 317) | 91.5 | 89.9 |
| Hospital (*nE* = 37; *nU* = 37) | 94.6 | 91.9 |
| Hospice (*nE* = 7; *nU* = 7) | 42.9 | 57.1 |
| HHA (*nE* = 255; *nU* = 255) | 95.3 | 93.3 |
| IRF (*nE* = 1; *nU* = 1) | 100 | 100 |
| LTCH (*nE* = 1; *nU* = 1) | 100 | 100 |
| SNF (*nE* = 104; *nU* = 104) | 86.5 | 90.4 |
| Other (*nE* = 20; *nU* = 20) | 80.0 | 95 |
| Clinical conditions (*nE* = 583; *nU* = 583) | | |
| Sepsis | | |
| Yes (*nE* = 25; *nU* = 25) | 96.0 | 88.0 |
| No (*nE* = 558; *nU* = 558) | 91.8 | 91.8 |
| Heart failure | | |
| Yes (*nE* = 132; *nU* = 132) | 93.9 | 94.7 |
| No (*nE* = 451; *nU* = 451) | 91.4 | 90.7 |
| Stroke | | |
| Yes (*nE* = 100; *nU* = 100) | 86.0 | 89.0 |
| No (*nE* = 483; *nU* = 483) | 93.2 | 92.1 |
| Hygiene—Toileting (*nE* = 567; *nU* = 567) | | |
| Independent (*nE* = 4; *nU* = 4) | 100 | 100 |
| Setup or clean-up assistance (*nE* = 22; *nU* = 22) | 100 | 95.5 |
| Supervision or touching assistance (*nE* = 122; *nU* = 122) | 91.0 | 90.2 |
| Partial/moderate assistance (*nE* = 136; *nU* = 136) | 91.9 | 91.2 |
| Substantial/maximal assistance (*nE* = 140; *nU* = 140) | 93.6 | 93.6 |
| Dependent (*nE* = 143; *nU* = 143) | 88.8 | 89.5 |
| Mobility—Transfer from lying to sitting (*nE* = 580[a]; *nU* = 580[a]) | | |
| Independent (*nE* = 41; *nU* = 41) | 95.1 | 97.6 |
| Setup or clean-up assistance (*nE* = 16; *nU* = 16) | 81.3 | 81.3 |
| Supervision or touching assistance (*nE* = 184; *nU* = 184) | 96.7 | 96.2 |
| Partial/moderate assistance (*nE* = 219; *nU* = 219) | 90.0 | 90.0 |
| Substantial/maximal assistance (*nE* = 93; *nU* = 93) | 91.4 | 88.2 |
| Dependent (*nE* = 27; *nU* = 27) | 77.8 | 81.5 |

NOTE: Because of differences in sample sizes for Expression and Understanding, we report sample sizes for each (*nE* = Expression; *nU* = Understanding).
[a] Significant ($p < 0.05$) associations with expression and understanding as indicated by chi-square tests of independence.

**Table A.27. Frequencies for Expression and Understanding Data Elements by Patient/Resident Characteristics and Clinical Groups in the LTCH Setting (percent)**

| Patient/Resident Characteristics and Clinical Groups | Expresses Without Difficulty (Yes) | Understands (Yes) |
|---|---|---|
| Gender (*nE* = 476; *nU* = 477) | | |
| Male (*nE* = 245; *nU* = 246) | 93.9 | 92.7 |
| Female (*nE* = 231; *nU* = 231) | 89.2 | 90.5 |
| Age (*nE* = 477; *nU* = 478[a]) | | |
| 18–44 (*nE* = 24; *nU* = 24) | 87.5 | 91.7 |
| 45–64 (*nE* = 118; *nU* = 119) | 93.2 | 95.0 |
| 65–74 (*nE* = 168; *nU* = 168) | 94.6 | 92.9 |
| 75–89 (*nE* = 152; *nU* = 152) | 88.2 | 89.5 |
| 90 or older (*nE* = 15; *nU* = 15) | 86.7 | 73.3 |
| Length of stay (*nE* = 420[a]; *nU* = 421[a]) | Yes: 23.3 (10.9)<br>No: 28.2 (12.4) | Yes: 23.4 (11.1)<br>No: 27.8 (12.0) |
| Disposition at discharge (*nE* = 460[a]; *nU* = 461) | | |
| Home (*nE* = 93; *nU* = 93) | 96.8 | 92.5 |
| Hospital (*nE* = 32; *nU* = 32) | 100 | 100 |
| Hospice (*nE* = 12; *nU* = 12) | 91.7 | 91.7 |
| HHA (*nE* = 77; *nU* = 77) | 97.4 | 97.4 |
| IRF (*nE* = 45; *nU* = 45) | 95.6 | 95.6 |
| LTCH (*nE* = 1; *nU* = 1) | 100 | 100 |
| SNF (*nE* = 134; *nU* = 134) | 83.6 | 86.6 |
| Other (*nE* = 66; *nU* = 67) | 84.9 | 86.6 |
| Clinical conditions (*nE* = 406; *nU* = 407) | | |
| Sepsis | | |
| Yes (*nE* = 67; *nU* = 68) | 91.0 | 91.2 |
| No (*nE* = 339; *nU* = 339) | 91.5 | 91.2 |
| Heart failure | | |
| Yes (*nE* = 14; *nU* = 14) | 92.9 | 92.9 |
| No (*nE* = 392; *nU* = 393) | 91.3 | 91.1 |
| Stroke | | |
| Yes (*nE* = 30; *nU* = 30[a]) | 70.0 | 76.7 |
| No (*nE* = 376; *nU* = 377) | 93.1 | 92.3 |

| Patient/Resident Characteristics and Clinical Groups | Expresses Without Difficulty (Yes) | Understands (Yes) |
|---|---|---|
| Hygiene—Toileting (*nE* = 392[a]; *nU* = 393[a]) | | |
| Independent (*nE* = 46; *nU* = 46) | 100 | 100 |
| Setup or clean-up assistance (*nE* = 33; *nU* = 33) | 93.9 | 90.9 |
| Supervision or touching assistance (*nE* = 59; *nU* = 59) | 96.6 | 95.0 |
| Partial/moderate assistance (*nE* = 51; *nU* = 51) | 94.1 | 98.0 |
| Substantial/maximal assistance (*nE* = 64; *nU* = 65) | 93.8 | 86.2 |
| Dependent (*nE* = 139; *nU* = 139) | 84.9 | 87.1 |
| Mobility—Transfer from lying to sitting (*nE* = 350[a]; *nU* = 350[a]) | | |
| Independent (*nE* = 64; *nU* = 64) | 98.4 | 98.4 |
| Setup or clean-up assistance (*nE* = 24; *nU* = 24) | 87.5 | 87.5 |
| Supervision or touching assistance (*nE* = 61; *nU* = 61) | 98.4 | 95.1 |
| Partial/moderate assistance (*nE* = 72; *nU* = 72) | 94.4 | 93.1 |
| Substantial/maximal assistance (*nE* = 50; *nU* = 50) | 94.0 | 90.0 |
| Dependent (*nE* = 79; *nU* = 79) | 83.5 | 83.5 |

NOTE: Because of differences in sample sizes for Expression and Understanding, we report sample sizes for each (*nE* = Expression; *nU* = Understanding).
[a] Significant ($p < 0.05$) associations with expression and understanding as indicated by chi-square tests of independence.

**Table A.28. Frequencies for Expression and Understanding Data Elements by Patient/Resident Characteristics and Clinical Groups in the SNF Setting (percent)**

| Patient/Resident Characteristics and Clinical Groups | Expresses Without Difficulty(Yes) | Understands(Yes) |
|---|---|---|
| Gender (*nE* = 1,107; *nU* = 1,104) | | |
| Male (*nE* = 431; *nU* = 430) | 90.5 | 88.1 |
| Female (*nE* = 676; *nU* = 674) | 92.8 | 91.1 |
| Age (*nE* = 1,102[a]; *nU* = 1,099[a]) | | |
| 18–44 (*nE* = 9; *nU* = 9) | 88.9 | 88.9 |
| 45–64 (*nE* = 75; *nU* = 75) | 98.7 | 98.7 |
| 65–74 (*nE* = 287; *nU* = 287) | 94.8 | 94.1 |
| 75–89 (*nE* = 556; *nU* = 553) | 89.8 | 87.9 |
| 90 or older (*nE* = 175; *nU* = 175) | 90.9 | 86.3 |
| Length of stay (*nE* = 956; *nU* = 953) | Yes: 21.4 (12.4)<br>No: 19.6 (10.5) | Yes: 21.3 (12.3)<br>No: 21.4 (12.6) |

| Patient/Resident Characteristics and Clinical Groups | Expresses Without Difficulty(Yes) | Understands(Yes) |
|---|---|---|
| Disposition at discharge (*nE* = 1,082[a]; *nU* = 1,079[a]) | | |
| Home (*nE* = 481; *nU* = 479) | 93.4 - | 92.3 |
| Hospital (*nE* = 111; *nU* = 111) | 86.5 | 84.7 |
| Hospice (*nE* = 10; *nU* = 9) | 100 | 100 |
| HHA (*nE* = 279; *nU* = 279) | 95.0 | 92.1 |
| IRF (*nE* = 1; *nU* = 1) | 100 | 100 |
| LTCH (*nE* = 10; *nU* = 10) | 80.0 | 100 |
| SNF (*nE* = 44; *nU* = 44) | 95.5 | 90.9 |
| Other (*nE* = 146; *nU* = 146) | 85.0 | 82.2 |
| Clinical conditions (*nE* = 867; *nU* = 864) | | |
| Sepsis | | |
| Yes (*nE* = 52; *nU* = 52) | 90.4 | 86.5 |
| No (*nE* = 815; *nU* = 812) | 91.9 | 90.2 |
| Heart failure | | |
| Yes (*nE* = 210; *nU* = 210) | 91.9 | 89.1 |
| No (*nE* = 657; *nU* = 654) | 91.8 | 90.2 |
| Stroke | | |
| Yes (*nE* = 66[a]; *nU* = 66[a]) | 80.3 | 78.8 |
| No (*nE* = 801; *nU* = 798) | 92.8 | 90.9 |
| Hygiene—Toileting (*nE* = 584; *nU* = 581[a]) | | |
| Independent (*nE* = 22; *nU* = 22) | 100 | 95.5 |
| Setup or clean-up assistance (nE = 24; *nU* = 24) | 95.8 | 100 |
| Supervision or touching assistance (*nE* = 143; *nU* = 141) | 93.0 | 93.6 |
| Partial/moderate assistance (*nE* = 182; *nU* = 180) | 91.8 | 90.6 |
| Substantial/maximal assistance (*nE* = 137; *nU* = 137) | 93.4 | 89.8 |
| Dependent (*nE* = 76; *nU* = 77) | 85.5 | 81.8 |
| Mobility—Transfer from lying to sitting (*nE* = 576; *nU* = 573) | | |
| Independent (*nE* = 58; *nU* = 57) | 94.8 | 93.0 |
| Setup or clean-up assistance (*nE* = 14; *nU* = 14) | 100 | 100 |
| Supervision or touching assistance (*nE* = 169; *nU* = 168) | 91.7 | 92.3 |
| Partial/moderate assistance (*nE* = 207; *nU* = 206) | 92.8 | 89.8 |
| Substantial/maximal assistance (*nE* = 101; *nU* = 101) | 92.1 | 89.1 |
| Dependent (*nE* = 27; *nU* = 27) | 81.5 | 81.5 |

NOTES: Because of differences in sample sizes for Expression and Understanding, we report sample sizes for each (*nE* = Expression; *nU* = Understanding).
[a] Significant ($p < 0.05$) associations with expression and understanding as indicated by chi-square tests of independence.

**Table A.29. Time to Complete the Expression and Understanding Data Elements by Urbanicity (minutes)**

| Market Group | Characteristic | Urban (A: *n* = 815; B: *n* = 735) | Nonurban (A: *n* = 0; B: *n* = 102) | Overall (A: *n* = 815; B: *n* = 837) |
|---|---|---|---|---|
| Market Group A (3–data element: a3, a4, a5) | Mean (SD) | 0.8 (0.4) | N/A | 0.8 (0.4) |
| Market Group B (2–data element: a6 and a7) | Mean (SD) | 0.7 (0.3) | 0.7 (0.3) | 0.7 (0.3) |

NOTE: N/A = not applicable.

**Table A.30. Time to Complete the Expression and Understanding Data Elements by Region (minutes)**

| Market Group | Characteristic | Northeast (A: *n* = 230; B: *n* = 222) | South (A: *n* = 264; B: *n* = 361) | Midwest (A: *n* = 134; B: *n* = 191) | West (A: *n* = 187; B: *n* = 63) | Overall (A: *n* = 815; B: *n* = 837) |
|---|---|---|---|---|---|---|
| Market Group A (3–data element: a3, a4, a5) | Mean (SD) | 0.8 (0.4) | 0.8 (0.4) | 0.8 (0.4) | 0.8 (0.4) | 0.8 (0.4) |
| Market Group B (2–data element: a6 and a7) | Mean (SD) | 0.7 (0.3) | 0.7 (0.3) | 0.7 (0.3) | 0.9 (0.4)[a] | 0.7 (0.3) |

[a] West region took significantly more time than other regions ($p < 0.001$).

**Table A.31. Time to Complete Expression and Understanding Section by Facility Ownership (minutes)**

| Market Group | Characteristic | For-Profit (A: *n* = 568; B: *n* = 450) | Nonprofit (A: *n* = 247; B: *n* = 376) | Overall (A: *n* = 815; B: *n* = 837) |
|---|---|---|---|---|
| Market Group A (three–data element: a3, a4, a5) | Mean (SD) | 0.8 (0.4) | 0.8 (0.4) | 0.8 (0.4) |
| Market Group B (two–data element: a6 and a7) | Mean (SD) | 0.7 (0.3) | 0.7 (0.3) | 0.7 (0.3) |

NOTE: Patient/resident numbers in for-profit and nonprofit categories in Market Group B do not sum to overall total because of missing profit status data.

**Table A.32. Time to Complete Expression and Understanding Section by Facility Size (minutes)**

| Market Group | Characteristic | Below Setting-Type Median (A: *n* = 293; B: *n* = 413) | Above Setting-Type Median (A: *n* = 522; B: *n* = 423) | Overall (A: *n* = 815; B: *n* = 837) |
|---|---|---|---|---|
| Market Group A (three–data element: a3, a4, a5) | Mean (SD) | 0.9 (0.4) | 0.8 (0.4) | 0.8 (0.4) |
| Market Group B (two–data element: a6 and a7) | Mean (SD) | 0.8 (0.4) | 0.6 (0.3) | 0.7 (0.3) |

NOTE: Patient/resident numbers in below and above setting-type median categories do not sum to overall total in Market Group B because of missing facility-size data.

**Table A.33. Interrater Reliability Kappa or Weighted Kappa for Expression and Understanding Data Elements by Urbanicity: Market Group A**

| Data Element | Urban (*n* = 441) | Nonurban (*n* = 0) |
|---|---|---|
| Speech clarity | 0.59 | — |
| Ability to express ideas and wants (a4) | 0.64 | — |
| Understanding verbal content (a5) | 0.59 | — |

NOTES: Interrater reliability not shown for data elements with proportions out of range for stable kappa estimate (per study power calculations). Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

**Table A.34. Interrater Reliability Kappa or Weighted Kappa for Expression and Understanding Data Elements by Urbanicity: Market Group B**

| Data Element | Urban (*n* = 451) | Nonurban (*n* = 67) |
|---|---|---|
| Express ideas and wants (a6) | 0.43 | 0.30 |
| Understands verbal content (a7) | 0.32 | 0.32 |

NOTE: Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

**Table A.35. Interrater Reliability Kappa or Weighted Kappa for Expression and Understanding Data Elements by Region: Market Group A**

| Data Element | Northeast (*n* = 73) | South (*n* = 158) | Midwest (*n* = 80) | West (*n* = 130) |
|---|---|---|---|---|
| Speech clarity | 0.49 | 0.45 | 0.65 | 0.85 |
| Ability to express ideas and wants (a4) | 0.68 | 0.60 | 0.46 | 0.82 |
| Understanding verbal content (a5) | 0.47 | 0.55 | 0.75 | 0.59 |

NOTE: Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

**Table A.36. Interrater Reliability Kappa or Weighted Kappa for Expression and Understanding Data Elements by Region: Market Group B**

| Data Element | Northeast (*n* = 39) | South (*n* = 200) | Midwest (*n* = 129) | West (*n* = 50) |
|---|---|---|---|---|
| Express ideas and wants (a6) | 0.33 | 0.25 | 0.73 | 0.58 |
| Understands verbal content (a7) | 0.40 | 0.13 | 0.47 | 0.27 |

NOTE: Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

**Table A.37. Interrater Reliability Kappa or Weighted Kappa for Expression and Understanding Data Elements by Facility Ownership: Market Group A**

| Data Element | For-Profit (*n* = 320) | Nonprofit (*n* = 121) |
|---|---|---|
| Speech clarity (a3) | — | — |
| Ability to express ideas and wants (a4) | 0.62 | — |
| Understanding verbal content (a5) | 0.55 | — |

NOTES: Interrater reliability not shown for data elements with proportions out of range for stable kappa estimate (per study power calculations). Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

**Table A.38. Interrater Reliability Kappa or Weighted Kappa for Expression and Understanding Data Elements by Facility Ownership: Market Group A**

| Data Element | For-Profit (*n* = 286) | Nonprofit (*n* = 226) |
|---|---|---|
| Expresses ideas and wants (a6) | 0.49 | 0.33 |
| Understands verbal content (a7) | 0.32 | 0.34 |

NOTE: Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

**Table A.39. Interrater Reliability Kappa or Weighted Kappa for Expression and Understanding Data Elements by Facility Size: Market Group A**

| Data Element | Below Setting-Type Median (*n* = 173) | Above Setting-Type Median (*n* = 268) |
|---|---|---|
| Speech clarity (a3) | — | 0.67 |
| Ability to express ideas and wants (a4) | 0.57 | 0.69 |
| Understanding verbal content (a5) | 0.61 | 0.57 |

NOTES: Interrater reliability not shown for data elements with proportions out of range for stable kappa estimate (per study power calculations). Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

**Table A.40. Interrater Reliability Kappa or Weighted Kappa for Expression and Understanding Data Elements by Facility Size: Market Group B**

| Data Element | Below Setting-Type Median (*n* = 251) | Above Setting-Type Median (*n* = 266) |
|---|---|---|
| Expresses ideas and wants (a6) | 0.42 | 0.42 |
| Understands verbal content (a7) | 0.29 | 0.35 |

NOTE: Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

# Supplementary Tables for Behavioral Signs and Symptoms

**Table A.41. Frequencies for Any Behavioral Symptom Exhibited by Patient/Resident Characteristics and Clinical Groups in the HHA Setting (percent)**

| Patient/Resident Characteristics and Clinical Groups | Any Behavioral Symptom Exhibited (Yes) |
|---|---|
| Gender (*n* = 611) | |
| Male (*n* = 216) | 2 |
| Female (*n* = 395) | 1 |
| Age (*n* = 608[a]) | |
| 18–44 (*n* = 4) | 25 |
| 45–64 (*n* = 60) | 5 |
| 65–74 (*n* = 173) | 1 |
| 75–89 (*n* = 303) | 1 |
| 90 or older (*n* = 68) | 3 |
| Length of stay (*n* = 489; mean, SD) | Yes: 28.25 (11.0)<br>No: 31.13 (15.67) |
| Disposition at discharge (*n* = 603) | |
| Home (*n* = 448) | 2 |
| Hospital (*n* = 23) | 0 |
| Hospice (*n* = 12) | 0 |
| HHA (*n* = 15) | 0 |
| IRF (*n* = 4) | 0 |
| LTCH (*n* = 1) | 0 |
| SNF (*n* = 6) | 0 |
| Other (*n* = 94) | 2 |
| Clinical conditions (*n* = 408) | |
| Sepsis | |
| Yes (*n* = 7) | 0 |
| No (*n* = 401) | 1 |
| Heart failure | |

| Patient/Resident Characteristics and Clinical Groups | Any Behavioral Symptom Exhibited (Yes) |
|---|---|
| Yes (*n* = 32) | 0 |
| No (*n* = 376) | 1 |
| Stroke | |
| Yes (*n* = 6) | 0 |
| No (*n* = 402) | 1 |
| Mobility—Transfer from lying to sitting (*n* = 381) | |
| Independent (*n* = 30) | 0 |
| Setup or clean-up assistance (*n* = 57) | 4 |
| Supervision or touching assistance (*n* = 116) | 1 |
| Partial/moderate assistance (*n* = 123) | 1 |
| Substantial/maximal assistance (*n* = 49) | 2 |
| Dependent (*n* = 6) | 0 |

[a] Significant ($p < 0.05$) associations with Any Behavioral Symptom Exhibited as indicated by chi-square tests of independence (ANOVA for length of stay).

**Table A.42. Frequencies for Any Behavioral Symptom Exhibited by Patient/Resident Characteristics and Clinical Groups in the IRF Setting (percent)**

| Patient/Resident Characteristics and Clinical Groups | Any Behavioral Symptom Exhibited(Yes) |
|---|---|
| Gender (*n* = 739) | |
| Male (*n* = 320) | 3 |
| Female (*n* = 419) | 2 |
| Age (*n* = 736) | |
| 18–44 (*n* = 5) | 0 |
| 45–64 (*n* = 58) | 3 |
| 65–74 (*n* = 287) | 1 |
| 75–89 (*n* = 333) | 2 |
| 90 or older (*n* = 53) | 8 |
| Length of stay (*n* = 721; mean, SD) | Yes: 14.13 (3.83)<br>No: 14.06 (5.08) |
| Disposition at discharge (*n* = 735) | |
| Home (*n* = 313) | 3 |
| Hospital (*n* = 37) | 0 |
| Hospice (*n* = 7) | 0 |
| HHA (*n* = 252) | 1 |
| IRF (*n* = 1) | 0 |
| LTCH (*n* = 0) | |
| SNF (*n* = 105) | 5 |
| Other (*n* = 20) | 0 |

| Patient/Resident Characteristics and Clinical Groups | Any Behavioral Symptom Exhibited(Yes) |
|---|---|
| Clinical conditions (*n* = 574) | |
| Sepsis | |
| Yes (*n* = 24) | 0 |
| No (*n* = 550) | 3 |
| Heart failure | |
| Yes (*n* = 128) | 1 |
| No (*n* = 446) | 3 |
| Stroke | |
| Yes (*n* = 99) | 0 |
| No (*n* = 475) | 3 |
| Hygiene—Toileting (*n* = 559) | |
| Independent (*n* = 5) | 0 |
| Setup or clean-up assistance (*n* = 21) | 0 |
| Supervision or touching assistance (*n* = 121) | 3 |
| Partial/moderate assistance (*n* = 136) | 1 |
| Substantial/maximal assistance (*n* = 138) | 4 |
| Dependent (*n* = 138) | 3 |
| Mobility—Transfer from lying to sitting (*n* = 571) | |
| Independent (*n* = 42) | 2 |
| Setup or clean-up assistance (*n* = 16) | 6 |
| Supervision or touching assistance (*n* = 185) | 3 |
| Partial/moderate assistance (*n* = 214) | 2 |
| Substantial/maximal assistance (*n* = 88) | 3 |
| Dependent (*n* = 26) | 4 |

**Table A.43. Frequencies for Any Behavioral Symptom Exhibited by Patient/Resident Characteristics and Clinical Groups in the LTCH Setting (percent)**

| Patient/Resident Characteristics and Clinical Groups | Any Behavioral Symptom Exhibited(Yes) |
|---|---|
| Gender (*n* = 452) | |
| Male (*n* = 231) | 2 |
| Female (*n* = 221) | 3 |
| Age (*n* = 453) | |
| 18–44 (*n* = 20) | 5 |
| 45–64 (*n* = 114) | 2 |
| 65–74 (*n* = 163) | 4 |
| 75–89 (*n* = 142) | 2 |

| Patient/Resident Characteristics and Clinical Groups | Any Behavioral Symptom Exhibited(Yes) |
|---|---|
| 90 or older (*n* = 14) | 0 |
| Length of stay (*n* = 400; mean, SD) | Yes: 25.90 (10.40)<br>No: 23.96 (11.12) |
| Disposition at discharge (*n* = 438) | |
| Home (*n* = 88) | 3 |
| Hospital (*n* = 30) | 0 |
| Hospice (*n* = 12) | 0 |
| HHA (*n* = 76) | 3 |
| IRF (*n* = 45) | 4 |
| LTCH (*n* = 1) | 0 |
| SNF (*n* = 124) | 3 |
| Other (*n* = 62) | 2 |
| Clinical conditions (*n* = 385) | |
| Sepsis | |
| Yes (*n* = 66) | 3 |
| No (*n* = 319) | 3 |
| Heart failure | |
| Yes (*n* = 13) | 0 |
| No (*n* = 372) | 3 |
| Stroke | |
| Yes (*n* = 27[a]) | 15 |
| No (*n* = 358) | 2 |
| Hygiene—Toileting (*n* = 373[a]) | |
| Independent (*n* = 45) | 2 |
| Setup or clean-up assistance (*n* = 33) | 0 |
| Supervision or touching assistance (*n* = 57) | 0 |
| Partial/moderate assistance (*n* = 49) | 0 |
| Substantial/maximal assistance (*n* = 60) | 10 |
| Dependent (*n* = 129) | 3 |
| Mobility—Transfer from lying to sitting (*n* = 335) | |
| Independent (*n* = 62) | 2 |
| Setup or clean-up assistance (*n* = 23) | 0 |
| Supervision or touching assistance (*n* = 58) | 0 |
| Partial/moderate assistance (*n* = 70) | 3 |
| Substantial/maximal assistance (*n* = 49) | 8 |
| Dependent (*n* = 73) | 1 |

[a] Significant ($p < 0.05$) associations with Any Behavioral Symptom Exhibited as indicated by chi-square tests of independence (ANOVA for length of stay).

**Table A.44. Frequencies for Any Behavioral Symptom Exhibited by Patient/Resident Characteristics and Clinical Groups in the SNF Setting (percent)**

| Patient/Resident Characteristics and Clinical Groups | Any Behavioral Symptom Exhibited(Yes) |
|---|---|
| Gender (*n* = 1,068) | |
| Male (*n* = 415) | 3 |
| Female (*n* = 653) | 1 |
| Age (*n* = 1,063) | |
| 18–44 (*n* = 9) | 0 |
| 45–64 (*n* = 72) | 1 |
| 65–74 (*n* = 280) | 1 |
| 75–89 (*n* = 533) | 2 |
| 90 or older (*n* = 169) | 2 |
| Length of stay (*n* = 927; mean, SD) | Yes: 22.53 (11.78)<br>No: 21.37 (12.22) |
| Disposition at discharge (*n* = 1,049[a]) | |
| Home (*n* = 466) | 2 |
| Hospital (*n* = 104) | 2 |
| Hospice (*n* = 9) | 11 |
| HHA (*n* = 274) | 0 |
| IRF (*n* = 1) | 0 |
| LTCH (*n* = 10) | 10 |
| SNF (*n* = 43) | 2 |
| Other (*n* = 142) | 3 |
| Clinical conditions (*n* = 837) | |
| Sepsis | |
| Yes (*n* = 51) | 0 |
| No (*n* = 786) | 2 |
| Heart failure | |
| Yes (*n* = 202) | 2 |
| No (*n* = 635) | 1 |
| Stroke | |
| Yes (*n* = 62) | 2 |
| No (*n* = 775) | 2 |

| Patient/Resident Characteristics and Clinical Groups | Any Behavioral Symptom Exhibited(Yes) |
|---|---|
| Hygiene—Toileting (*n* = 563) | |
| Independent (*n* = 22) | 5 |
| Setup or clean-up assistance (*n* = 23) | 0 |
| Supervision or touching assistance (*n* = 140) | 2 |
| Partial/moderate assistance (*n* = 172) | 2 |
| Substantial/maximal assistance (*n* = 132) | 2 |
| Dependent (*n* = 74) | 3 |
| Mobility—Transfer from lying to sitting (*n* = 555) | |
| Independent (*n* = 58) | 0 |
| Setup or clean-up assistance (*n* = 12) | 0 |
| Supervision or touching assistance (*n* = 163) | 2 |
| Partial/moderate assistance (*n* = 202) | 1 |
| Substantial/maximal assistance (*n* = 95) | 4 |
| Dependent (*n* = 25) | 0 |

[a] Significant ($p < 0.05$) associations with Any Behavioral Symptom Exhibited as indicated by chi-square tests of independence (ANOVA for length of stay).

**Table A.45. Time to Complete the Behavioral Signs and Symptoms Data Elements by Urbanicity (minutes)**

| Characteristic | Urban (*n* = 1,439) | Nonurban (*n* = 104) | Overall (*n* = 1,543) |
|---|---|---|---|
| Mean (SD) | 1.4 (0.7) | 1.7 (0.8) | 1.4 (0.8) |

**Table A.46. Time to Complete the Behavioral Signs and Symptoms Data Elements by Region (minutes)**

| Characteristic | Northeast (*n* = 408) | South (*n* = 578) | Midwest (*n* = 315) | West (*n* = 242) | Overall (*n* = 1,543) |
|---|---|---|---|---|---|
| Mean (SD) | 1.6 (0.7) | 1.4 (0.7) | 1.3 (0.8) | 1.6 (0.8) | 1.4 (0.8) |

**Table A.47. Time to Complete the Behavioral Signs and Symptoms Data Elements by Facility Ownership (minutes)**

| Characteristic | For-Profit (*n* = 959) | Nonprofit (*n* = 571) | Overall (*n* = 1,543) |
|---|---|---|---|
| Mean (SD) | 1.4 (0.7) | 1.5 (0.8) | 1.4 (0.8) |

NOTE: Patient/resident numbers in for-profit and nonprofit categories do not sum to overall total because of missing profit status data.

**Table A.48. Time to Complete the Behavioral Signs and Symptoms Data Elements by Facility Size (minutes)**

| Characteristic | Below Setting-Type Median (*n* = 669) | Above Setting-Type Median (*n* = 873) | Overall (*n* = 1,543) |
|---|---|---|---|
| Mean (SD) | 1.5 (0.7) | 1.4 (0.8) | 1.4 (0.8) |

NOTE: Patient/resident numbers in below and above setting-type median categories do not sum to overall total because of missing facility-size data.

# References


Arciniegas, David B., Kerri Held, and Peter Wagner, "Cognitive Impairment Following Traumatic Brain Injury," *Current Treatment Options in Neurology*, Vol. 4, No. 1, February 2002, pp. 43–57.

Beck, Cornelia, Linda Rossby, and Beverly Baldwin, "Correlates of Disruptive Behavior in Cognitively Impaired Elderly Nursing Home Residents," *Archives of Psychiatric Nursing*, Vol. 5, No. 5, October 1991, pp. 281–291.

Beck, Cornelia, Lori Frank, Neale R. Chumbler, Patricia O'Sullivan, Therese S. Vogelpohl, Joyce Rasin, Robert Walls, and Beverly Baldwin, "Correlates of Disruptive Behavior in Severely Cognitively Impaired Nursing Home Residents," *The Gerontologist*, Vol. 38, No. 2, April 1998, pp. 189–198.

Benjamini, Yoav, and Yosef Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 57, No. 1, 1995, pp. 289–300.

Borthwick, Sheena, "Communication Impairment in Patients Following Stroke," *Nursing Standard*, Vol. 26, No. 19, January 2012, pp. 35–41.

Buslovich, Steven, and Gary J. Kennedy, "Potential Effect of Screening for Subtle Cognitive Deficits on Hospital Readmission," *Journal of the American Geriatrics Society*, Vol. 60, No. 10, October 2012, pp. 1980–1981.

CAM—*See* Sharon K. Inouye, Christopher H. van Dyck, Cathy A. Alessi, Sharyl Balkin, Alan P. Siegal, and Ralph I. Horwitz, "Clarifying Confusion: The Confusion Assessment Method: A New Method for Detection of Delirium," *Annals of Internal Medicine*, Vol. 113, No. 12, December 1990, pp. 941–948.

Cameron, Jill I., Angela M. Cheung, David L. Streiner, Peter C. Coyte, and Donna E. Stewart, "Stroke Survivors' Behavioral and Psychologic Symptoms Are Associated with Informal Caregivers' Experiences of Depression," *Archives of Physical Medicine and Rehabilitation*, Vol. 87, No. 2, February 2006, pp. 177–183.

Campbell, Noll L., Malaz A. Boustani, Elaine N. Skopelja, Sujuan Gao, Fred W. Unverzagt, and Michael D. Murray, "Medication Adherence in Older Adults with Cognitive Impairment: A Systematic Evidence-Based Review," *American Journal of Geriatric Pharmacotherapy*, Vol. 10, No. 3, June 2012, pp. 165–177.

Campbell, Susan E., D. Gwyn Seymour, William R. Primrose, Joanna E. Lynch, Edmund Dunstan, Mireia Espallargues, Giovanni Lamura, Peter Lawson, Ian Philp, Elizabeth

Mestheneos, Barbara Politynska, Ismo Raiha, and the ACME*plus* Project Team, "A Multi-Centre European Study of Factors Affecting the Discharge Destination of Older People Admitted to Hospital: Analysis of In-Hospital Data from the ACME*plus* Project," *Age and Ageing*, Vol. 34, No. 5, September 2005, pp. 467–475.

Carpenter, G. Iain, Charlotte L. Hastie, John N. Morris, Brant E. Fries, and Joel Ankri, "Measuring Change in Activities of Daily Living in Nursing Home Residents with Moderate to Severe Cognitive Impairment," *BMC Geriatrics*, Vol. 6, No. 7, 2006.

Centers for Medicare & Medicaid Services, *Development of Functional Outcome Quality Measures for Skilled Nursing Facilities (SNFs): Public Comment Summary Report*, Baltimore, Md., September 2016.

———, "Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long-Term Care Hospital Prospective Payment System and Proposed Policy Changes and Fiscal Year 2018 Rates; Quality Reporting Requirements for Specific Providers; Medicare and Medicaid Electronic Health Record (EHR) Incentive Program Requirements for Eligible Hospitals, Critical Access Hospitals, and Eligible Professionals; Provider-Based Status of Indian Health Service and Tribal Facilities and Organizations; Costs Reporting and Provider Requirements; Agreement Termination Notices," *Federal Register*, Vol. 82, No. 81, April 28, 2017a, pp. 19796–20231.

———, "Inpatient Rehabilitation Facility Prospective Payment System for Federal Fiscal Year 2018," *Federal Register*, Vol. 82, No. 84, May 3, 2017b, pp. 20690–20747.

———, "Prospective Payment System and Consolidated Billing for Skilled Nursing Facilities for FY 2018, SNF Value-Based Purchasing Program, SNF Quality Reporting Program, Survey Team Composition, and Proposal To Correct the Performance Period for the NHSN HCP Influenza Vaccination Immunization Reporting Measure in the ESRD QIP for PY 2020," *Federal Register*, Vol. 82, No. 85, May 4, 2017c, pp. 21014–21100.

———, "CY 2018 Home Health Prospective Payment System Rate Update and Proposed CY 2019 Case-Mix Adjustment Methodology Refinements; Home Health Value-Based Purchasing Model; and Home Health Quality Reporting Requirements," *Federal Register*, Vol. 82, No. 144, July 28, 2017d, pp. 35270–35393.

———, *Development and Maintenance of Post-Acute Care Cross-Setting Standardized Assessment Data: Public Comment Summary Report 2*, Baltimore, Md., January 2018.

Chodosh, Joshua, Maria Orlando Edelen, Joan L. Buchanan, Julia Ann Yosef, Joseph G. Ouslander, Dan R. Berlowitz, Joel E. Streim, and Debra Saliba, "Nursing Home Assessment of Cognitive Impairment: Development and Testing of a Brief Instrument of Mental Status," *Journal of the American Geriatrics Society*, Vol. 56, No. 11, November 2008, pp. 2069–2075.

CMS—*See* Centers for Medicare & Medicaid Services.

Cohen, Jacob, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed., New York: Routledge, 2013.

Cruz-Oliver, Dulce M., T. K. Malmstrom, C. M. Allen, N. Tumosa, and J. E. Morley, "The Veterans Affairs Saint Louis University Mental Status Exam (SLUMS Exam) and the Mini-Mental Status Exam as Predictors of Mortality and Institutionalization," *Journal of Nutrition, Health, and Aging*, Vol. 16, No. 7, August 2012, pp. 636–641.

Diwan, Sadhna, and Victoria L. Phillips, "Agitation and Dementia-Related Problem Behaviors and Case Management in Long-Term Care," *International Psychogeriatrics*, Vol. 13, No. 1, March 2001, pp. 5–21.

Edelen, Maria Orlando, Barbara J. Gage, Adam J. Rose, Sangeeta C. Ahluwalia, Amy Soo Jin DeSantis, Michael Stephen Dunbar, Shira H. Fischer, Wenjing Huang, David J. Klein, Steven Martino, Francesca Pillemer, Tepring Piquado, Victoria Shier, Regina A. Shih, Cathy D. Sherbourne, and Brian D. Stucky, *Development and Maintenance of Standardized Cross Setting Patient Assessment Data for Post-Acute Care: Summary Report of Findings from Alpha 1 Pilot Testing*, Santa Monica, Calif.: RAND Corporation, RR-1895-CMS, 2017. As of June 4, 2019: https://www.rand.org/pubs/research_reports/RR1895.html

Edelen, Maria Orlando, Anthony Rodriguez, Sangeeta C. Ahluwalia, Emily K. Chen, Sarah Dalton, Michael Stephen Dunbar, Jason Michel Etchegaray, Shira H. Fischer, Wenjing Huang, David J. Klein, Jaime Madrigano, Monique Martineau, Steven Martino, Patrick Orr, Jessica Phillips, Francesca Pillemer, Debra Saliba, Shoshana R. Shelton, Cathy D. Sherbourne, Victoria Shier, and Regina A. Shih, *Development and Maintenance of Standardized Cross Setting Patient Assessment Data for Post-Acute Care: Summary Report of Findings from Alpha 2 Pilot Testing*, Baltimore, Md.: Centers for Medicare & Medicaid Services, June 2018. As of June 4, 2019: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/Alpha-2-SPADE-Pilot-Summary-Document.pdf

Edelen, Maria Orlando, Anthony Rodriguez, Sangeeta C. Ahluwalia, Emily K. Chen, Sarah Dalton, Michael Stephen Dunbar, Shira H. Fischer, David J. Klein, Monique Martineau, Patrick Orr, Jessica Phillips, Debra Saliba, Victoria Shier, and Regina A. Shih, *Development and Evaluation of Candidate Standardized Patient Assessment Data Elements: Findings from the National Beta Test (Volume 2: Background and Methods)*, Santa Monica, Calif.: RAND Corporation, RR-3004-CMS, 2019a. As of August 20, 2019: https://www.rand.org/pubs/research_reports/RR3004.html

Edelen, Maria Orlando, Anthony Rodriguez, Sangeeta C. Ahluwalia, Emily K. Chen, Catherine C. Cohen, Sarah Dalton, Jason Michel Etchegaray, Wenjing Huang, Brenda Karkos, David J. Klein, Jaime Madrigano, Monique Martineau, Terry Moore, Patrick Orr, Jessica Phillips, Ben Schalet, Victoria Shier, Susan M. Paddock, and Denis Agniel, *Development and Evaluation of Candidate Standardized Patient Assessment Data Elements: Findings from the National Beta Test (Volume 3: Sample Description)*, Santa Monica, Calif.: RAND Corporation, RR-3004/2-CMS, 2019b. As of August 20, 2019: https://www.rand.org/pubs/research_reports/RR3004z2.html

Edlund, Agneta, Maria Lundstrom, Benny Brännström, Gösta Bucht, and Yngve Gustafson, "Delirium Before and After Operation for Femoral Neck Fracture," *Journal of the American Geriatrics Society*, Vol. 49, No. 10, October 2001, pp. 1335–1340.

Feil, Denis, Tonya Marmon, and Jürgen Unützer, "Cognitive Impairment, Chronic Medical Illness, and Risk of Mortality in an Elderly Cohort," *American Journal of Geriatric Psychiatry*, Vol. 11, No. 5, September–October 2003, pp. 551–560.

Gage, Barbara, Melvin J. Ingber, Melissa Morley, Laura Smith, Anne Deutsch, Tracy Kline, Jill Dever, Judith Abbate, Richard Miller, Brieanne Lyda-McDonald, Cynthia Kelleher, Danielle Garfinkel, Joshua Manning, Christopher M. Murtaugh, Margaret Stineman, and Trudy Mallinson, *Post-Acute Care Payment Reform Demonstration: Final Report*, Research Triangle Park, N.C.: RTI International, March 2012.

Givens, Jane L., Tara B. Sanft, and Edward R. Marcantonio, "Functional Recovery After Hip Fracture: The Combined Effects of Depressive Symptoms, Cognitive Impairment, and Delirium," *Journal of the American Geriatrics Society*, Vol. 56, No. 6, June 2008, pp. 1075–1079.

Gray, Shelly L., Jane E. Mahoney, and David K. Blough, "Adverse Drug Events in Elderly Patients Receiving Home Health Services Following Hospital Discharge," *Annals of Pharmacotherapy*, Vol. 33, No. 11, 1999, pp. 1147–1153.

Hayes, Tamara L., Nicole Larimer, Andre Adami, and Jeffrey A. Kaye, "Medication Adherence in Healthy Elders: Small Cognitive Changes Make a Big Difference," *Journal of Aging and Health*, Vol. 21, No. 4, June 2009, pp. 567–580.

Heruti, Raphael J., Ayala Lusky, Rachel Dankner, Haim Ring, Mark Dolgopiat, Vita Barell, Shalom Levenkrohn, and Abraham Adunsky, "Rehabilitation Outcome of Elderly Patients After a First Stroke: Effect of Cognitive Status at Admission on the Functional Outcome," *Archives of Physical Medicine and Rehabilitation*, Vol. 83, No. 6, June 2002, pp. 742–749.

Horowitz, Amy, "Vision Impairment and Functional Disability Among Nursing Home Residents," *The Gerontologist*, Vol. 34, No. 3, June 1994, pp. 316–323.

Hugo, Julie, and Mary Ganguli, "Dementia and Cognitive Impairment: Epidemiology, Diagnosis, and Treatment," *Clinics in Geriatric Medicine*, Vol. 30, No. 3, August 2014, pp. 421–442.

Inouye, Sharon K., "Delirium in Older Persons," *New England Journal of Medicine*, Vol. 354, No. 11, March 16, 2006, pp. 1157–1165.

Inouye, Sharon K., Christopher H. van Dyck, Cathy A. Alessi, Sharyl Balkin, Alan P. Siegal, and Ralph I. Horwitz, "Clarifying Confusion: The Confusion Assessment Method: A New Method for Detection of Delirium," *Annals of Internal Medicine*, Vol. 113, No. 12, December 1990, pp. 941–948.

Kaufer, Daniel I., Jeffrey L. Cummings, Dianne Christine, Tim Bray, Steve Castellon, Donna Masterman, Audrey MacMillan, Patrick Ketchel, and Steven T. DeKosky, "Assessing the Impact of Neuropsychiatric Symptoms in Alzheimer's Disease: The Neuropsychiatric Inventory Caregiver Distress Scale," *Journal of the American Geriatrics Society*, Vol. 46, No. 2, February 1998, pp. 210–215.

Kiely, Dan K., Margaret A. Bergmann, Richard N. Jones, Katharine M. Murphy, E. John Orav, and Edward R. Marcantonio, "Characteristics Associated with Delirium Persistence Among Newly Admitted Post-Acute Facility Patients," *Journals of Gerontology: Series A (Biological Sciences and Medical Sciences)*, Vol. 59, No. 4, April 2004, pp. 344–349.

Kim, J. S., Suhin Choi, Sun U. Kwon, and Y. S. Seo, "Inability to Control Anger or Aggression After Stroke," *Neurology*, Vol. 58, No. 7, 2002, pp. 1106–1108.

Lee, Hochang B., Simon C. Mears, Paul B. Rosenberg, Jeannie-Marie S. Leoutsakos, Allan Gottschalk, and Frederick E. Sieber, "Predisposing Factors for Postoperative Delirium After Hip Fracture Repair in Individuals With and Without Dementia," *Journal of the American Geriatrics Society*, Vol. 59, No. 12, December 2011, pp. 2306–2313.

Logsdon, Rebecca G., Laura E. Gibbons, Susan M. McCurry, and Linda Teri, "Assessing Quality of Life in Older Adults with Cognitive Impairment," *Psychosomatic Medicine*, Vol. 64, No. 3, May 2002, pp. 510–519.

Lorig, Kate, David S. Sobel, Philip L. Ritter, Dennis Laurent, and Mitchell Hobbs, "Effect of a Self-Management Program on Patients with Chronic Disease," *Effective Clinical Practice*, Vol. 4, No. 6, November–December 2001, pp. 256–262.

Lundstrom, Maria, Birgitta Olofsson, Michael Stenvall, Stig Karlsson, Lars Nyberg, Undis Englund, Bengt Borssen, Olle Svensson, and Yngve Gustafson, "Postoperative Delirium in Old Patients with Femoral Neck Fracture: A Randomized Intervention Study," *Aging Clinical and Experimental Research*, Vol. 19, No. 3, June 2007, pp. 178–186.

Mantel, Nathan, and William Haenszel, "Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease," *Journal of the National Cancer Institute*, Vol. 22, No. 4, April 1959, pp. 719–748.

Marcantonio, Edward R., Lee Goldman, Carol M. Mangione, Lynn E. Ludwig, Brenda Muraca, Christine M. Haslauer, Magruder C. Donaldson, Anthony D. Whittemore, David J. Sugarbaker, Robert Poss, Susan Haas, E. Francis Cook, John Orav, and Thomas H. Lee, "A Clinical Prediction Rule for Delirium After Elective Noncardiac Surgery," *JAMA*, Vol. 271, No. 2, January 12, 1994, pp. 134–139.

Marcantonio, Edward R., Dan K. Kiely, Samuel E. Simon, E. John Orav, Richard N. Jones, Katharine M. Murphy, and Margaret A. Bergmann, "Outcomes of Older People Admitted to Postacute Facilities with Delirium," *Journal of the American Geriatrics Society*, Vol. 53, No. 6, June 2005, pp. 963–969.

Marcantonio, Edward R., Samuel E. Simon, Margaret A. Bergmann, Richard N. Jones, Katharine M. Murphy, and John N. Morris, "Delirium Symptoms in Post-Acute Care: Prevalent, Persistent, and Associated with Poor Functional Recovery," *Journal of the American Geriatrics Society*, Vol. 51, No. 1, January 2003, pp. 4–9.

Mehta, Sangeeta, Deborah Cook, John Devlin, Yoanna Skrobik, Maureen Meade, Dean Fergusson, Margaret Herridge, Marilyn Steinberg, John Granton, Niall Ferguson, Maged Tanios, Peter Dodek, Robert Fowler, Karen Burns, Michael Jacka, Kendiss Olafson, Ranjeeta Mallick, Steven Reynolds, Sean Keenan, and Lisa Burry, "Prevalence, Risk Factors, and Outcomes of Delirium in Mechanically Ventilated Adults," *Critical Care Medicine*, Vol. 43, No. 3, March 2015, pp. 557–566.

Millar, Keith, A. J. Asbury, and Gordon D. Murray, "Pre-Existing Cognitive Impairment as a Factor Influencing Outcome After Cardiac Surgery," *British Journal of Anaesthesia*, Vol. 86, No. 1, January 2001, pp. 63–67.

Oh, Esther S., Frederick E. Sieber, Jeannie-Marie Leoutsakos, Sharon K. Inouye, and Hochang B. Lee, "Sex Differences in Hip Fracture Surgery: Preoperative Risk Factors for Delirium and Postoperative Outcomes," *Journal of the American Geriatrics Society*, Vol. 64, No. 8, August 2016, pp. 1616–1621.

Pandharipande, Pratik, Bryan A. Cotton, Ayumi Shintani, Jennifer Thompson, Brenda Truman Pun, John A. Morris, Jr., Robert Dittus, and E. Wesley Ely, "Prevalence and Risk Factors for Development of Delirium in Surgical and Trauma Intensive Care Unit Patients," *Journal of Trauma*, Vol. 65, No. 1, July 2008, pp. 34–41.

RAND Corporation, *Technical Expert Panel Summary/Expert Input Report: Development and Maintenance of Post-Acute Care Cross-Setting Standardized Patient Assessment Data*, Santa

Monica, Calif., WR-1187-CMMS, 2017a. As of June 28, 2019: https://www.rand.org/pubs/working_papers/WR1187.html

———, *Technical Expert Panel Summary (Second Convening): Development and Maintenance of Post-Acute Care Cross-Setting Standardized Patient Assessment Data*, Santa Monica, Calif., WR-1184-CMMS, 2017b. As of June 28, 2019: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/RAND-IMPACT-TEP-Second-Convening-Final-Report-March-2017.pdf

Resnick, Helaine E., Brant E. Fries, and Lois M. Verbrugge, "Windows to Their World: The Effect of Sensory Impairments on Social Engagement and Activity Time in Nursing Home Residents," *Journals of Gerontology: Series B (Social Sciences)*, Vol. 52B, No. 3, May 1997, pp. S135–S144.

Rock, P. L. , J. P. Roiser, W. J. Riedel, and A. D. Blackwell, "Cognitive Impairment in Depression: A Systematic Review and Meta-Analysis," *Psychological Medicine*, Vol. 44, No. 10, July 2004, pp. 2029–2040.

Saliba, Debra, Joan Buchanan, Maria Orlando Edelen, Joel Streim, Joseph Ouslander, Dan Berlowitz, and Joshua Chodosh, "MDS 3.0: Brief Interview for Mental Status," *Journal of the American Medical Directors Association*, Vol. 13, No. 7, September 2012, pp. 611–617.

Shi, Qiyun, Roseanna Presutti, Daniel Selchen, and Gustavo Saposnik, "Delirium in Acute Stroke: A Systematic Review and Meta-Analysis," *Stroke*, Vol. 43, No. 3, March 2012, pp. 645–649.

Stuck, Andreas E., Jutta M. Walthert, Thorsten Nikolaus, Christophe J. Büla, Christoph Hohmann, and John C. Beck, "Risk Factors for Functional Status Decline in Community-Living Elderly People: A Systematic Literature Review," *Social Science & Medicine*, Vol. 48, No. 4, February 1999, pp. 445–469.

Sun, Jia-Hao, Lan Tan, and Jin-Tai Yu, "Post-Stroke Cognitive Impairment: Epidemiology, Mechanisms and Management," *Annals of Translational Medicine*, Vol. 2, No. 8, August 2014, p. 80.

Tan, Lay Ling, Hwee Bee Wong, and Harry Allen, "The Impact of Neuropsychiatric Symptoms of Dementia on Distress in Family and Professional Caregivers in Singapore," *International Psychogeriatrics*, Vol. 17, No. 2, June 2005, pp. 253–263.

Unverzagt, Frederick W., Sujuan Gao, Olusegun Baiyewu, Adesola Ogunniyi, Oye Gureje, Anthony Perkins, Christine L. Emsley, Jeanne Dickens, Rebecca Margrete Evans, Beverly Musick, Kathryn Sandra Kaur. Hall, Siu L. Hui, and Hugh C. Hendrie, "Prevalence of Cognitive Impairment: Data from the Indianapolis Study of Health and Aging," *Neurology*, Vol. 57, No. 9, November 13, 2001, pp. 1655–1662.

U.S. Department of Agriculture, "Documentation: 2010 Rural–Urban Commuting Area (RUCA) Codes," webpage, last updated October 12, 2016. As of June 4, 2019: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/documentation/

Vogels, Raymond L. C., Philip Scheltens, Jutta M. Schroeder-Tanka, and Henry C. Weinstein, "Cognitive Impairment in Heart Failure: A Systematic Review of the Literature," *European Journal of Heart Failure*, Vol. 9, No. 5, May 2007, pp. 440–449.

Voyer, Philippe, René Verreault, Ginette M. Azizah, Johanne Desrosiers, Nathalie Champoux, and Annick Bédard, "Prevalence of Physical and Verbal Aggressive Behaviours and Associated Factors Among Older Adults in Long-Term Care Facilities," *BMC Geriatrics*, Vol. 5, No. 13, 2005.