Development and Evaluation of Candidate Standardized Patient Assessment Data Elements: Findings from the National Beta Test (Volume 8: Observational Assessments of Cognitive Function, Mental Status, and Pain)

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-3004/7-CMS

Development and Evaluation of Candidate Standardized Patient Assessment Data Elements:

Findings from the National Beta Test

(Volume 8: Observational Assessments of Cognitive Function, Mental Status, and Pain)

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/RR3004Z7

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

Tables

Abbreviations

ADLs	activities of daily living
BIMS	Brief Interview for Mental Status
CARE	Continuity Assessment Record and Evaluation
CMS	Centers for Medicare & Medicaid Services
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
OASIS	Outcome and Assessment Information Set
PAC	post-acute care
PAC-PRD	Post-Acute Care Payment Reform Demonstration
PHQ	Patient Health Questionnaire
PHQ-9-OV	Patient Health Questionnaire (Observational Version)
SD	standard deviation
SPADE	standardized patient assessment data element
SNF	skilled nursing facility
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 8 of the final report on the National Beta Test, which includes the identification and testing of candidate SPADEs developed specifically for patients/residents who are unable to communicate (staff assessments of mental status, mood, and pain). 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 this 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/Calendar Year 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, being 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 completed by research nurse and facility/agency staff assessor pairs to allow for evaluation of

[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.

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 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 set for all 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 characteristic 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.

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]),[7] 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 later in this volume within the specific data element chapter for which a difference emerged.

[7] U.S. Department of Agriculture, 2016.

Statistical Tests

- Categorical associations were statistically evaluated using chi-square tests of independence and, in the case of ordinal data, Mantel-Haenszel chi-square.[8] 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 one continuous and one categorical variable 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.[9] 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 corrected significance levels using the Benjamini-Hochberg method, where each significance test is evaluated against an adjusted critical value.[10] We set our desired level of significance at 0.01 to minimize Type I error and increase confidence in significant effects.

[8] Mantel and Haenszel, 1959.

[9] Cohen, 2013.

[10] Benjamini and Hochberg, 1995.

2. Observational Assessments of Cognitive Function, Mental Status, and Pain

Data elements under development for the assessment of Cognitive Function, Mental Status, and Pain (see Volumes 4 and 5) are conducted through patient/resident interviews. Not all patients/residents receiving PAC services are able to participate in these interviews because of cognitive impairment, for example, or difficulty communicating. Yet it is still important to assess these clinical categories among this *non-communicative* population. This volume covers data elements that assess for Cognitive Function (i.e., cognitive impairment), Mental Status (i.e., depression), and Pain through observation rather than interviews. The second convening of the TEP considered observation-based data elements in each of these categories. This chapter briefly reviews stakeholder feedback from the TEP and other sources on those data elements in the stages before the National Beta Test.

Cognitive Function

As described in Volume 4, cognitive impairment is associated with several disorders, conditions, and injuries,[11] as well as functional limitations in physical ability,[12] social relationships,[13] the ability to adhere to health care regimens,[14] and decisionmaking.[15] Impaired cognitive function is also associated with an increased likelihood of hospital readmission following discharge to PAC.[16] Because patients/residents in PAC settings are at risk for cognitive impairment, it is important to assess cognitive function to screen for impairment, assess the severity of a disorder, monitor the progression of symptoms, and develop and maintain an appropriate care plan. Patients/residents who are unable to complete the BIMS, either because verbal or nonverbal responses cannot be understood or because the patient refuses to continue (even if cognitively intact), would be eligible for the observation-based assessment described in this section.

[11] Rock et al., 2004; Hugo and Ganguli, 2014; Sun, Tan, and Yu, 2014; Arciniegas, Held, and Wagner, 2002.

[12] Rosano et al., 2005; Wang et al., 2002.

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

[14] Campbell et al., 2012.

[15] Kim, Karlawish, and Caine, 2002.

[16] Gage et al., 2012.

Information Gathering

The MDS includes an observation-based assessment of cognitive function called the Staff Assessment of Mental Status. This data element is an observational assessment of long-term memory, short-term memory, memory or recall ability, and decisionmaking, based on staff observation and intended for use among patients/residents who are unable to communicate. Studies testing this data element in nursing home patients/residents have shown it to have good interrater reliability ($r = 0.80$)[17] and good validity based on its correlation with other assessments, such as the Blessed Test ($r = 0.66$, $p < 0.05$) and the Reisberg Global Deterioration Scale ($r = 0.59$, $p < 0.05$).[18] During validation testing of the MDS in nursing home residents, the Staff Assessment of Mental Status demonstrated substantial to almost perfect agreement across all items (interrater reliability ranging from 0.80 to 0.90).[19] A shortened version of the Staff Assessment of Mental Status was also tested in the Post-Acute Care Payment Reform Demonstration (PAC-PRD). In that demonstration, the shortened Staff Assessment of Mental Status showed no discordant assessment pairs for patients' ability to recall the current season, the location of their room, and staff names and faces. For the other items, interrater reliability ranged from 0.58 to 0.88.[20] Although we did identify alternative assessments of cognitive function that could be used in this population (e.g., the Blessed Test), the strong performance data and demonstrated cross-setting feasibility of the Staff Assessment of Mental Status led us to pursue this assessment over others.

Stakeholder Feedback and Field Testing

TEP members at the second convening, during which these data elements were first put forth for discussion, raised several concerns with using the Staff Assessment of Mental Status across all PAC settings, including phrasing of some of the data elements and instructions. One member questioned whether it would be better to have assessors alter their administration of the BIMS, perhaps with the use of pictures, instead of using a different assessment altogether that is less sensitive. Another asserted that capturing and addressing the reason for the inability to communicate should be of greater concern than administering a separate data element set.

The Staff Assessment of Mental Status data elements were also submitted for public comment in 2017. Commenters noted that the Staff Assessment of Mental Status is important because its completion can help identify significant issues among patients/residents who are unable to communicate that can affect their health and plan of care, as well as transitions among providers. Other commenters agreed that the assessment seems valid and/or reliable and that it

[17] Casten et al., 1998.

[18] Lawton et al., 1998.

[19] Saliba and Buchanan, 2008.

[20] Gage et al., 2012.

could be helpful and potentially capable of filling a void in describing case mix associated with patients/residents who are unable to complete the BIMS. On the other hand, some commenters questioned the data element set's impact on improving quality, and others did not see its value in describing case mix or raised concerns about cross-setting applicability.

Testing of the Staff Assessment of Mental Status data element set in the Alpha 2 pilot test found that interrater agreement tended to be high for the components of this data element set, except in the IRF setting for the question regarding knowledge of staff names and faces. On average, assessments took longer to complete for research nurses than facility staff and took longer among patients in LTCHs than in other settings. In some cases, assessors indicated confusion on some wording and an inability to complete portions of the data element set because of missing documentation.

Candidate SPADEs in the National Beta Test

With consensus from stakeholders that an observation-based assessment of cognitive function would be consistent with the intent of the Improving Medicare Post-Acute Care Transformation (IMPACT) Act, and in consideration of the results from Alpha 2 testing, the Staff Assessment of Mental Status data element set was tested in the National Beta Test. Results are described in Chapter 3.

Mental Status

As described in Volume 5, depression is a common mental health problem in older adults and is particularly common in PAC settings.[21] Depression can negatively affect many aspects of health and well-being, including quality of life,[22] physical function,[23] pain,[24] rejection of care behaviors,[25] and increased mortality from other causes.[26] Depression screeners help PAC providers better understand the needs of their patients and residents by prompting further evaluation and, after establishing an appropriate diagnosis related to depressive symptoms, elucidating the patient's or resident's ability to participate in therapies (e.g., physical rehabilitation) during his or her stay, as well as identifying appropriate ongoing treatment and support needs at the time of discharge. Patients/residents who are unable to complete the

[21] Ell et al., 2006; Hyer, 2005; Blazer, 2002; Jones, Marcantonio, and Rabinowitz, 2003; Payne et al., 2002; Teresi et al., 2001.

[22] Diefenbach, Tolin, and Gilliam, 2012; Garrison, Overcash, and McMillan, 2011; Heisel et al., 2010; Kroenke et al., 2010; Ruo et al., 2003.

[23] Slaughter et al., 2011.

[24] Lapane et al., 2012; Leone, Standoli, and Hirth, 2009.

[25] Ishii, Streim, and Saliba, 2010; Ishii, Streim, and Saliba, 2012.

[26] Charney et al., 2003; Harris and Cooper, 2006; Kane, Yochim, and Lichtenberg, 2010; Ziegelstein, 2001.

interview-based PHQ,[27] in either its nine- or two-question form, to assess the presence and frequency of depressive symptoms would be eligible for the observation-based assessment described in this section.

Information Gathering

The PHQ-9-OV (Observational Version) is a version of the PHQ-9 that includes all nine questions in the PHQ-9 plus a question assessing the symptom of temperament (which is readily observed and can be indicative of depression among this population) and is administered through observation. It is included in the MDS, has been validated in the nursing home population, and has demonstrated feasibility in that setting.[28] Because of its demonstrated validity and feasibility for use among nursing home residents and its correspondence with the PHQ-9, this assessment was considered for cross-setting standardized assessment of depressed mood via staff observation.

Stakeholder Feedback and Field Testing

During the second convening of the group, TEP members offered generally positive feedback about the PHQ-9-OV. Suggestions for improvement included minor modifications and changes to descriptions of the assessment process in training materials and user guides. TEP members did not express great concern over the issue of burden for this data element.

In the second public comment period (2017), several commenters supported the PHQ-9-OV and shared that the data element set could improve the quality and appropriateness of services for patients/residents who have difficulty communicating. Comments also touched on the benefits of the level of detail the data element set could offer and noted that the PHQ-9-OV could identify important health issues, ultimately improving a patient's/resident's care planning and transitions of care communication. Moreover, commenters added that the PHQ-9-OV appears valid, reliable, feasible, and potentially able to fill a void in describing case mix. Conversely, several commenters were concerned about burden, the data element set's recollection time period of two weeks, validity, and cross-setting relevance.

The PHQ-9-OV data element set was included in the Alpha 2 pilot test. Interrater reliability was high for all data elements and varied little across settings. No data elements were skipped incorrectly or completed incorrectly, there were almost no missing data, and the time required to complete the assessment was reasonable. Among LTCHs, there was a high percentage of "unknown/unable to assess" codes, suggesting that the PHQ-9-OV may be less feasible to administer in LTCHs than in other PAC settings. Additionally, the small number of non-

[27] Spitzer, Kroenke, and Williams, 1999. The Patient Health Questionnaire (PHQ) was developed by Pfizer Inc. © 1999 Pfizer Inc. All rights reserved.

[28] Saliba et al., 2012.

communicative HHA patients included in this field test makes it difficult to draw a conclusion about feasibility in that setting.

Candidate SPADEs in the National Beta Test

Given the feedback from stakeholders and testing results, the PHQ-9-OV was included in the National Beta Test. Results are described in Chapter 4.

Pain

As described in Volume 5, pain is a common condition among adults of all ages.[29] Regular pain occurs in 25 percent to 80 percent of residents in SNFs and other institutional care settings, such as nursing homes, making it more common than many other chronic conditions and symptoms.[30] Pain is frequent among those receiving home health care, with 53 percent reporting daily pain interfering with activity on admission.[31] Pain in older adults occurs in conjunction with many acute and chronic conditions, such as osteoarthritis, leg pain during the night, cancer and cancer treatment, and peripheral vascular disease.[32] Conditions causing pain in older adults may be associated with depression,[33] sleep disturbance,[34] and reduced participation in rehabilitation activities.[35]

Volume 5 describes multiple interview-based data elements for assessment of pain that were tested in the National Beta Test; patients/residents who are unable to complete those data elements would be eligible for the observation-based data element set described in this section.

Information Gathering

Observational data elements to assess pain are included in the OASIS and the MDS. The OASIS queries the frequency of pain interfering with a patient's/resident's activity or movement. The set of data elements included in the MDS document indicators of pain or possible pain across four types of behaviors, including nonverbal sounds, vocal complaints of pain, facial expressions, and protective body movements or postures (e.g., clutching or holding a body part during movement). This data element set has demonstrated excellent reliability, with kappas of 0.94 and 0.96. A nearly identical version of the MDS observational pain assessment was tested in

[29] Dahlhamer et al., 2018.

[30] Abdulla et al., 2013; Shen et al., 2015.

[31] Murtaugh et al., 2008.

[32] American Geriatrics Society Panel on the Pharmacological Management of Persistent Pain in Older Persons, 2009.

[33] Sullivan-Singh et al., 2014.

[34] Blytt et al., 2018; Eslami et al., 2016.

[35] Brenner and Marsella, 2008; Chin, Ho, and Cheung, 2013; Zanca et al., 2013.

the PAC-PRD and demonstrated adequate reliability for nonverbal sounds, vocal complaints of pain, and facial expression (kappas ranging from 0.61 to 0.66) but lower agreement for protective body movements or postures (kappa of 0.42).

A web-based search and literature review to identify candidate data elements for assessing pain in non-communicative patients/residents yielded 39 additional data elements, many of which are similar in terms of content. The American Geriatrics Society has developed guidelines for pain assessment among individuals who are unable to communicate, and the first three categories of behaviors (facial expressions, verbalizations and vocalizations, and body movements) are typically covered in all observational pain assessments. Notable assessments from research and the literature review included the Abbey Pain Scale, the Checklist of Nonverbal Pain Indicators, the CAN Pain Assessment Tool, the DOLOPLUS-2, the Mahoney Pain Scale, the Pain Assessment Checklist for Seniors with Limited Ability to Communicate, and the Pain Assessment in Advanced Dementia.

Observational assessments for pain in the literature evaluate a variety of overlapping but not completely correspondent sets of pain indicators and use several scoring algorithms. Our clinical advisers recommended adopting a routine assessment that indicates behaviors present both at rest and during activity. They supported cross-setting standardization of the data element tested in the PAC-PRD but suggested modifications, including clarifying the instructions to encourage assessment during daily care activities (when indicators of pain are most likely to be observed) and enhancing the verbal descriptions for each assessed behavior to better align with published guidelines.

Stakeholder Feedback and Field Testing

Based on the results of our information-gathering activities and in consideration of the feedback from our clinical advisers, we presented a modified version of the data element set in use in the current MDS and tested in the PAC-PRD to federal subject-matter experts. The subject-matter experts recommended adding two items to assess (1) the frequency with which observed indicators of pain were observed and (2) whether observed indicators of pain resolved or diminished in response to administration of pain medications or treatments.

The members of our TEP agreed that observational assessments for pain are important to offer for patients/residents who cannot complete interview-based assessments, and the changes proposed by our clinical advisers and the subject-matter experts were well received. Much of the discussion centered around the instructions for when to complete the assessment and the need to standardize the conditions in which the observational assessment was conducted without increasing assessor burden. Overall, TEP members rated the observational pain data element set the highest among the observational data elements.

The Observational Assessment of Pain or Distress data element set was presented for public comment in 2017. Although there was general support for this data element set, one commenter

suggested that some of the listed behaviors are not specifically indicative for pain and recommended further evaluation.

The Observational Assessment of Pain or Distress data element set was included in the Alpha 2 pilot test. Assessors considered evidence of behavioral indicators of pain when the patient/resident was engaged in care activities (e.g., during transfer procedures, repositioning, bathing, range of motion or other exercises), when indicators are most likely to be observed. Multiple sources of information were used to complete these items during the three-day assessment time frame, including direct patient/resident observation, medical records, and feedback from the patient's/resident's direct care staff.

Interrater agreement was substantial to almost perfect (kappas ranging from 0.69 to 1.00), and the data element set took approximately three to five minutes to complete. Assessor feedback reflected that the pain items were straightforward but somewhat challenging to complete because of the time required for observation and the need to consult multiple data sources. Overall, the pain items were reliable and feasible to administer and did not appear to require any further changes.

Candidate SPADEs in the National Beta Test

The Observational Assessment of Pain or Distress was included in the National Beta Test, given the feedback from stakeholders and testing results. Results are described in Chapter 5.

Summary of Candidate SPADEs in the National Beta Test

The Staff Assessment of Mental Status, PHQ-9-OV, and Observational Assessment of Pain or Distress data elements that were evaluated in the National Beta Test are shown in Table 2.1. This table also lists the evaluative and input opportunities in which each data element has been included during the contract period, specific National Beta Test design features relevant to the data element, and an indication of its use in any of the four PAC assessments.

Table 2.1. Data Elements Evaluated in the National Beta Test Non-Communicative Sample

Data Element	Input Opportunities	National Beta Test Inclusion Notes	Current Assessment Instrument Use
Staff Assessment of Mental Status	Alpha 2, Public Comment 2	For patients/residents unable to communicate	MDS
Staff Assessment of Patient/Resident Mood (PHQ-9-OV)	Alpha 2, Public Comment 2	For patients/residents unable to communicate	MDS
Observational Assessment of Pain or Distress	Alpha 2, Public Comment 2	For patients/residents unable to communicate	MDS

3. Observational Assessment of Cognitive Function

Data Element Description

The Staff Assessment of Mental Status is an observational assessment of cognitive function that assesses long-term memory, short-term memory, memory or recall ability, and decisionmaking, based on staff observation; information provided by staff, family, and friends; and medical records. The assessment, shown in Figure 3.1, includes data elements documenting short- and long-term memory, recall ability, and cognitive skills for daily decisionmaking and is intended for use among patients/residents in all PAC settings who were unable to complete the interview-administered BIMS because of nonsensical answers or an inability to make themselves understood at least some of the time. As described in Volume 4 (Cognitive Function), cognitive impairment has been linked to limitations in the capacity to make informed decisions about health care[36] and adhere to medication regimens,[37] a lower quality of life,[38] decreased social functioning, decreased ability to maintain personal relationships,[39] and decreased functional status.[40] Conducting cognitive assessments is critically important to screen for cognitive impairment, rate severity of disorder, develop a care plan, and monitor progression.

The Staff Assessment of Mental Status data elements are completed through observation of the patient/resident, communication with staff and other caregivers, and review of the patient's/resident's medical record. The Staff Assessment of Mental Status is currently collected in the MDS and IRF-PAI.

[36] Lorig et al., 2001.

[37] Campbell et al., 2012.

[38] Logsdon et al., 2002.

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

[40] Campbell et al., 2005; Stuck et al., 1999.

Figure 3.1. Staff Assessment of Mental Status Data Elements

B3a. Short-term Memory OK: Seems or appears to recall after 5 minutes

- ☐ 0 = Memory OK
- ☐ 1 = Memory problem
- ☐ 9 = **Unknown or unable to assess**

B3b. Long-term Memory OK: Seems or appears to recall long past

- ☐ 0 = Memory OK
- ☐ 1 = Memory problem
- ☐ 9 = **Unknown or unable to assess**

B3c. Memory/Recall Ability: Is the patient/resident normally able to recall:

B3c1. Current season

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

B3c2. Location of own room

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

B3c3. Staff names and faces

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

B3c4. That he or she is in a nursing facility/hospital bed/rehabilitation facility/home

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

B3d. Cognitive Skills for Daily Decision Making: Made decisions regarding tasks of daily life:

- ☐ 0 = Independent – decisions consistent/reasonable
- ☐ 1 = Modified independence – some difficulty in new situations only
- ☐ 2 = Moderately impaired – decisions poor; cues/supervision required
- ☐ 3 = Severely impaired – never/rarely made decisions
- ☐ 9 = **Unknown or unable to assess**

Testing Objectives

As described in Volume 2, the non-communicative assessments were administered at a single point not directly tied to an admission or discharge to maximize the number of patients/residents eligible for these assessments during the National Beta Test field period. Basic descriptive statistics (e.g., frequencies, means, SDs) are presented for the Staff Assessment of Mental Status data elements to characterize the rates of impairment for patients/residents in each setting and for the overall sample. We also examined the data element "ability to make decisions about everyday tasks" by patient/resident characteristics and clinical groups of interest. Feasibility (rates of missingness and time to complete) and interrater reliability (kappa and percent agreement) were examined as well.

Results

Feasibility

Frequencies/Missing

Table 3.1 shows the percentage of responses for each Staff Assessment of Mental Status data element overall and by setting. The Staff Assessment of Mental Status was administered to 513 of the 548 eligible patients/residents, or 93.6 percent: 32 in HHAs, 103 in IRFs, 169 in LTCHs, and 209 in SNFs. Among these, overall missing data at the data element level ranged from 2.9 percent to 33.5 percent, the majority of which reflected cases in which responses were unknown or unable to be assessed . One data element ("ability to make decisions regarding everyday tasks") had only 2.9 percent missing, perhaps because it is the most easily observed.. However, the missing data ranged from 23.4 percent to 33.5 percent among the other data elements. At the setting level, missing data rates were similar; however, they were slightly higher in LTCHs (HHAs: 6.3 percent to 21.9 percent, IRFs: 2.9 percent to 30.1 percent, LTCHs: 3.6 percent to 55.0 percent, SNFs: 2.9 percent to 33.5 percent). Overall, the Staff Assessment of Mental Status results show that 84 percent of non-communicative patients/residents had a short-term memory problem and 76 percent had a long-term memory problem. As expected, setting type was associated with impairment $\left(\chi^2_{(9)} = 30.16, p < 0.001\right)$, such that a higher percentage of patients/residents who were severely impaired in the ability to make everyday decisions were found in LTCH and SNF settings (82 percent and 81 percent, respectively), compared with 67 percent and 57 percent of patients/residents in HHAs and IRFs, respectively. All data elements tended to follow this trend, in which scores reflected a less impaired cognitive status for patients/residents in HHA and IRF settings than in LTCH and SNF settings.

Table 3.1. Overall and Setting-Specific Response Frequencies for Staff Assessment of Mental Status Data Elements Among Non-Communicative Patients/Residents (percent)

Data Element	HHA (*n* = 32)	IRF (*n* = 103)	LTCH (*n* = 169)	SNF (*n* = 209)	Overall (*n* = 513)
Short-term memory OK (b3a)					
Memory OK	25	20	24	8	16
Memory problem	75	80	76	92	84
Long-term memory OK (b3b)					
Memory OK	54	32	24	16	24
Memory problem	46	68	76	84	76
Is the patient normally able to recall: current season (b3c1)					
Yes	20	38	12	5	15
Is the patient normally able to recall: location of own room (b3c2)					
Yes	50	20	7	17	17
Is the patient normally able to recall: staff names and faces (b3c3)					
Yes	46	50	33	37	39
Is the patient normally able to recall: that he or she is in a care facility (b3c4)					
Yes	63	63	41	22	39
Ability to make decisions regarding everyday tasks (b3d)					
Independent	0	1	1	1	1
Modified independence	7	7	3	1	3
Moderately impaired	27	35	15	16	20
Severely impaired	67	57	82	81	76

Exploratory Comparisons with Known Groups

In other volumes of this report, we present associations between patient/resident performance on the candidate SPADEs and other known patient/resident characteristics. In those cases, we formed expectations about the associations based on the available research literature on populations who are similar to the Beta sample (e.g., patients/residents receiving PAC, older adults, nursing home residents). Observing expected or logical associations contributed to evidence that the data elements are valid—that is, that they assess the construct that they are intended to capture. However, our ability to generate hypotheses for the non-communicative sample was limited, primarily because of the lack of research on equivalent populations (i.e., patients/residents receiving PAC services who are unable to communicate by any means). In addition, the smaller sample of non-communicative patients/residents limits the power of these analyses to detect differences between rates of less common conditions or characteristics, such as

sepsis or certain discharge disposition categories. For these reasons, we consider the comparisons with known groups described below to be exploratory. We present them as part of this volume for completeness but not as evidence for or against the construct validity of the candidate SPADEs.

Table 3.2 shows rates of patients/residents characterized as *severely impaired* on the "ability to make decisions regarding everyday tasks" data element for the full non-communicative sample (all settings combined), 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 or older), 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).

Table 3.2. Frequencies for Ability to Make Decisions Regarding Everyday Tasks by Patient/Resident Characteristics and Clinical Groups (percent)

Patient/Resident Characteristics and Clinical Groups	Severely Impaired
Gender (*n* = 490)	
Male (*n* = 221)	72.8
Female (*n* = 269)	78.1
Age (*n* = 488)	
18–44 (*n* = 27)	81.5
45–64 (*n* = 95)	68.4
65–74 (*n* = 100)	78.0
75–89 (*n* = 192)	76.6
90 or older (*n* = 74)	77.0
Length of stay (*n* = 257;[a] mean, SD)	Yes: 26.5 (14.0) No: 22.8 (11.9)
Disposition at discharge (*n* = 476)[a]	
Home (*n* = 53)	58.5
Hospital (*n* = 46)	76.1
Hospice (*n* = 16)	81.3
HHA (*n* = 30)	50.0
IRF (*n* = 17)	64.7
SNF (*n* = 143)	78.3
LTCH (*n* = 0)	0.0
Other (*n* = 170)	83.0

Patient/Resident Characteristics and Clinical Groups	Severely Impaired
Clinical conditions (*n* = 240)	
Sepsis	
Yes (*n* = 23)	75.5
No (*n* = 217)	73.4
Heart failure	
Yes (*n* = 224)	59.1
No (*n* = 16)	74.9
Stroke	
Yes (*n* = 47)	66.7
No (*n* = 193)	76.2
Hygiene—Toileting (*n* = 203)[a]	
Independent (*n* = 2)	50.0
Setup or clean-up assistance (*n* = 1)	0.0
Supervision or touching assistance (*n* = 5)	60.0
Partial/moderate assistance (*n* = 11)	27.3
Substantial/maximal assistance (*n* = 27)	48.2
Dependent (*n* = 157)	79.0
Mobility—Transfer from lying to sitting (*n* = 178)[a]	
Independent (*n* = 4)	0.0
Setup or clean-up assistance (*n* = 1)	0.0
Supervision or touching assistance (*n* = 15)	60.0
Partial/moderate assistance (*n* = 26)	42.3
Substantial/maximal assistance (*n* = 36)	63.9
Dependent (*n* = 96)	78.1

[a] Significant ($p < 0.05$) associations with impaired "ability to make decisions regarding everyday tasks" as indicated by chi-square tests of independence.

Because of the heterogeneous nature of the non-communicative population and lack of research on equivalent populations, we did not form hypotheses or expectations about the associations among patient/resident characteristics.

For gender, age, and all clinical conditions, there were no significant associations with patient/resident impaired ability to make decisions regarding everyday tasks. However, impaired ability to make decisions was associated with length of stay, disposition at discharge, and ADLs.

Length of Stay and Disposition at Discharge:

- There was a significant association between length of stay and impaired ability to make decisions ($F_{(1,255)} = 4.2$, $p < 0.05$), such that length of stay was longer on average for patients/residents who were severely impaired (mean [M] = 26.5, SD = 14.0) compared with patients/residents who were not severely impaired (M = 22.8, SD = 11.9). *Although we did not have expectations about any associations, this association aligns with the idea*

that patients/residents with more-severe cognitive impairment may require longer periods of rehabilitation or nursing care, either because of their cognitive abilities or underlying conditions that may be affecting their cognition.

- Disposition at discharge was significantly associated with an impaired ability to make decisions $\left(\chi^2_{(6)} = 26.01, p < 0.01\right)$, such that patients/residents with severely impaired ability to make decisions were discharged at higher rates to hospices, hospitals, SNFs, or "other" placements (e.g., group homes, assisted living facilities) relative to rates of discharge to LTCHs, HHAs, IRFs, and home. Rates of discharge to hospitals were roughly equal to overall rates of severe impairment in the ability to make everyday decisions (76 percent). *Although we did not have expectations about any associations, this conforms with a scenario of more-impaired patients/residents being discharged to relatively higher-intensity care settings.*

ADLs: Toileting and Ability to Transfer from Lying to Sitting

- Impaired ability to make decisions was also associated with independence levels on both toileting hygiene $\left(\chi^2_{(5)} = 25.06, p < 0.01\right)$ and ability to transfer from lying to sitting $\left(\chi^2_{(5)} = 22.90, p < 0.01\right)$, such that severely impaired patients/residents were more likely to be completely dependent or require substantial or maximal assistance. *Although we did not have expectations about any associations, this finding is logically consistent considering the likely complexity and personal assistance needs of these patients/residents.*

Time to Complete

Table 3.3 shows the average time to complete the Staff Assessment of Mental Status data elements overall and by setting. On average, it took 2.6 minutes (SD = 1.6) to complete overall and ranged from 2.4 minutes (SD = 1.5) in IRFs to 3.4 minutes (SD = 2.3) in HHAs. There were no statistically significant differences among settings on time to complete the Staff Assessment of Mental Status.

Table 3.3. Time to Complete the Staff Assessment of Mental Status Data Elements (minutes)

Time to Complete	HHA (*n* = 28)	IRF (*n* = 99)	LTCH (*n* = 148)	SNF (*n* = 214)	Overall (*n* = 489)
Mean (SD)	3.4 (2.3)	2.4 (1.5)	2.6 (1.4)	2.5 (1.6)	2.6 (1.6)

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.1–A.4 in the appendix). The Staff Assessment of Mental Status data elements took significantly less time to complete in the Midwest region (2.3 minutes [SD = 1.5]) than in the South (2.7 minutes [SD = 1.7], Cohen's d = 0.25) and in the West (2.7 minutes [SD = 1.5], Cohen's d = 0.27). There were no other significant differences in

time to complete the Staff Assessment of Mental Status data elements in these sensitivity analyses.

Interrater Reliability

Table 3.4 shows the kappa interrater reliability coefficients for the Staff Assessment of Mental Status overall and by setting. Kappa coefficients were computed on 505 patients/residents for whom paired observational assessments were completed. Overall kappa coefficients ranged from 0.74 to 0.94, with all but one (0.74 for the short-term memory data element) falling in the excellent classification range. The overall kappa for this data element was good (0.74) but slightly lower than other data elements. At the setting level, kappa coefficients were good to excellent and ranged from 0.79 to 1.00 in HHAs, 0.64 to 0.87 in IRFs, 0.74 to 0.94 in LTCHs, and 0.72 to 0.98 in SNFs.

Table 3.4. Interrater Reliability Kappa or Weighted Kappa for Staff Assessment of Mental Status Data Elements

Data Element	HHA (*n* = 32)	IRF (*n* = 101)	LTCH (*n* = 165)	SNF (*n* = 207)	Overall (*n* = 505)
Short-term memory OK (b3a)	0.82	0.64	0.77	0.72	0.74
Long-term memory OK (b3b)	0.84	0.85	0.74	0.82	0.82
Is the patient normally able to recall: current season (b3c1)	0.86	0.84	0.89	—	0.90
Is the patient normally able to recall: location of own room (b3c2)	0.93	0.87	—	0.98	0.94
Is the patient normally able to recall: staff names and faces (b3c3)	0.79	0.84	0.94	0.90	0.89
Is the patient normally able to recall: that he or she is in a care facility (b3c4)	0.83	0.80	0.92	0.91	0.89
Ability to make decisions regarding everyday tasks (b3d)	1.00	0.78	0.92	0.92	0.89

NOTES: Interrater reliability 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, and 0.81–1.00 is excellent/almost perfect.

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.5–A.8 in the appendix). The Northeast region had a lower kappa for the recall of current season data element (0.55) relative to the other regions (0.94 to 1.00) and substantially lower kappa for the short-term memory data element (0.39) relative to the other regions (0.72 to 0.83). Additionally, the kappa for the short-term memory data element was substantially lower for nonprofit facilities (0.30) compared with

for-profit facilities (0.86). No other noteworthy differences were found for interrater reliability of the Staff Assessment of Mental Status data elements in these sensitivity analyses.

Table 3.5 shows percent agreement for Staff Assessment of Mental Status data elements overall and by setting. Overall percent agreement was high for all data elements, ranging from 93 percent to 98 percent, with minimal setting differences. At the setting level, percent agreement was high for all data elements and ranged from 89 percent to 100 percent in HHs, 88 percent to 96 percent in IRFs, 90 percent to 98 percent in LTCHs, and 95 percent to 100 percent in SNFs.

Table 3.5. Interrater Reliability—Percent Agreement for Staff Assessment of Mental Status Data Elements

Data Element	HHA (*n* = 32)	IRF (*n* = 101)	LTCH (*n* = 165)	SNF (*n* = 207)	Overall (*n* = 505)
Short-term memory OK (b3a)	93	89	91	95	93
Long-term memory OK (b3b)	92	93	90	95	93
Is the patient normally able to recall: current season (b3c1)	96	93	98	100	98
Is the patient normally able to recall: location of own room (b3c2)	96	96	98	99	98
Is the patient normally able to recall: staff names and faces (b3c3)	89	92	97	95	95
Is the patient normally able to recall: that he or she is in a care facility (b3c4)	92	90	96	97	95
Ability to make decisions regarding everyday tasks (b3d)	100	88	97	98	96

Assessor Feedback

According to the assessor survey, facility/agency staff and research nurses found the Staff Assessment of Mental Status to be somewhat to moderately clinically useful. When asked about burden in the assessor survey, facility/agency staff on average thought that it was "slightly difficult" to collect information, and the data element scored in the middle range on burden relative to other data elements. In contrast, research nurses rated this data element as one of the most burdensome to collect, most likely because of their limited familiarity and contact with the patients/residents.

Summary

Results for the Staff Assessment of Mental Status indicate moderate overall support for cross-setting standardization. Assessors considered Staff Assessment of Mental Status data elements to be somewhat to moderately clinically useful but to have moderate to high data collection burden compared with the other data elements. Although only a few associations between an impaired ability to make decisions and patient/resident characteristics were observed,

the associations aligned with our expectation that patients/residents with more-severe cognitive impairment typically need more time for rehabilitation, more intensive care, and more assistance with ADLs.

In terms of feasibility, the rates of missing data were high for most of the data elements in this assessment. The missing data are due to the staff observer's inability to determine an accurate answer based on his or her observation of the patient/resident and are an inherent challenge with all observational assessments with non-communicative patients/residents. Time to complete was under three minutes, with no differences across settings. However, the data elements were completed more quickly in the Midwest relative to other regions. These effects were small but exceeded our cutoff effect size value of 0.2, raising some questions about the generalizability of the time-to-complete estimate. Psychometric performance of the data elements was quite good: Kappas for the Staff Assessment of Mental Status were good to excellent and percent agreement was high for all data elements. These combined results show moderate feasibility, good to excellent interrater reliability, moderate clinical utility, and moderate data collection burden for the Staff Assessment of Mental Status as a candidate data element for standardization across PAC settings.

4. Staff Assessment of Patient/Resident Mood (PHQ-9-OV)

Data Element Description

The PHQ-9-OV[41] assesses signs and symptoms of depressed mood in patients/residents who cannot complete a patient/resident mood interview because of an inability to communicate. As described in Chapter 3 of Volume 5 (Mental Status and Pain), screening for signs and symptoms of depression is important because undetected depression can lead to degraded physical and mental health and functioning, [42] increased medical care utilization and costs,[43] reduced quality of life,[44] and premature death.[45]

The PHQ-9-OV data elements are completed through interviews with staff, family members, and/or other caregivers who know the patient/resident best and by reviewing medical records. The PHQ-9-OV is currently used in the MDS. These data elements are shown in Figure 4.1.

Figure 4.1. Staff Assessment of Patient/Resident Mood (PHQ-9-OV) Data Elements

E4a1. SYMPTOM PRESENCE: Over the last 2 weeks, did the patient/resident have little interest or pleasure in doing things?

- ☐ 0 = No **[SKIP TO E4b1]**
- ☐ 1 = Yes
- ☐ 9 = **Unknown or unable to assess [SKIP TO E4b1]**

E4a2. SYMPTOM FREQUENCY: Over the last 2 weeks, how often did the patient/resident have little interest or pleasure in doing things?

- ☐ 0 = Never or 1 day
- ☐ 1 = 2-6 days (several days)
- ☐ 2 = 7-11 days (half or more of the days)
- ☐ 3 = 12-14 days (nearly every day)
- ☐ 9 = **Unknown or unable to assess**

[41] Spitzer, Kroenke, and Williams, 1999. The Patient Health Questionnaire (PHQ) was developed by Pfizer Inc. © 1999 Pfizer Inc. All rights reserved.

[42] Cronin-Stubbs et al., 2000.

[43] Katon et al., 2003.

[44] Diefenbach, Tolin, and Gilliam, 2012.

[45] Lépine and Briley, 2011; Schoevers et al., 2000.

E4b1. SYMPTOM PRESENCE: Over the last 2 weeks, did the patient/resident feel or appear down, depressed, or hopeless?

- ☐ 0 = No **[SKIP TO E4c1]**
- ☐ 1 = Yes
- ☐ 9 = **Unknown or unable to assess [SKIP TO E4c1]**

E4b2. SYMPTOM FREQUENCY: Over the last 2 weeks, how often did the patient/resident feel or appear down, depressed, or hopeless?

- ☐ 0 = Never or 1 day
- ☐ 1 = 2-6 days (several days)
- ☐ 2 = 7-11 days (half or more of the days)
- ☐ 3 = 12-14 days (nearly every day)
- ☐ 9 = **Unknown or unable to assess**

E4c1. SYMPTOM PRESENCE: Over the last 2 weeks, did the patient/resident have trouble falling or staying asleep, or sleeping too much?

- ☐ 0 = No **[Skip to E4d1]**
- ☐ 1 = Yes
- ☐ 9 = **Unknown or unable to assess [Skip to E4d1]**

E4c2. SYMPTOM FREQUENCY: Over the last 2 weeks, how often did the patient/resident have trouble falling or staying asleep, or sleeping too much?

- ☐ 0 = Never or 1 day
- ☐ 1 = 2-6 days (several days)
- ☐ 2 = 7-11 days (half or more of the days)
- ☐ 3 = 12-14 days (nearly every day)
- ☐ 9 = **Unknown or unable to assess**

E4d1. SYMPTOM PRESENCE: Over the last 2 weeks, did the patient/resident feel tired or have little energy?

- ☐ 0 = No **[SKIP to E4e1]**
- ☐ 1 = Yes
- ☐ 9 = **Unknown or unable to assess [SKIP to E4e1]**

E4d2. SYMPTOM FREQUENCY: Over the last 2 weeks, how often did the patient/resident feel tired or have little energy?

- ☐ 0 = Never or 1 day
- ☐ 1 = 2-6 days (several days)
- ☐ 2 = 7-11 days (half or more of the days)
- ☐ 3 = 12-14 days (nearly every day)
- ☐ 9 = **Unknown or unable to assess**

E4e1. SYMPTOM PRESENCE: Over the last 2 weeks, did the patient/resident have a poor appetite or overeating?

- ☐ 0 = No **[SKIP TO E4f1]**
- ☐ 1 = Yes
- ☐ 9 = **Unknown or unable to assess [SKIP TO E4f1]**

E4e2. SYMPTOM FREQUENCY: Over the last 2 weeks, how often did the patient/resident have a poor appetite or overeating?

- ☐ 0 = Never or 1 day
- ☐ 1 = 2-6 days (several days)
- ☐ 2 = 7-11 days (half or more of the days)
- ☐ 3 = 12-14 days (nearly every day)
- ☐ 9 = **Unknown or unable to assess**

E4f1. SYMPTOM PRESENCE: Over the last 2 weeks, did the patient/resident indicate that s/he feels bad about self, is a failure, or has let self or family down?

- ☐ 0 = No **[SKIP TO E4g1]**
- ☐ 1 = Yes
- ☐ 9 = **Unknown or unable to assess [SKIP TO E4g1]**

E4f2. SYMPTOM FREQUENCY: Over the last 2 weeks, how often did the patient/resident indicate that s/he feels bad about self, is a failure, or has let self or family down?

- ☐ 0 = Never or 1 day
- ☐ 1 = 2-6 days (several days)
- ☐ 2 = 7-11 days (half or more of the days)
- ☐ 3 = 12-14 days (nearly every day)
- ☐ 9 = **Unknown or unable to assess**

E4g1. SYMPTOM PRESENCE: Over the last 2 weeks, did the patient/resident have trouble concentrating on things, such as reading the newspaper or watching television?

- ☐ 0 = No **[SKIP TO E4h1]**
- ☐ 1 = Yes
- ☐ 9 = **Unknown or unable to assess [SKIP TO E4h1]**

E4g2. SYMPTOM FREQUENCY: Over the last 2 weeks, how often did the patient/resident have trouble concentrating on things, such as reading the newspaper or watching television?

- ☐ 0 = Never or 1 day
- ☐ 1 = 2-6 days (several days)
- ☐ 2 = 7-11 days (half or more of the days)
- ☐ 3 = 12-14 days (nearly every day)
- ☐ 9 = **Unknown or unable to assess**

E4h1. SYMPTOM PRESENCE: Over the last 2 weeks, did the patient/resident move or speak so slowly that other people have noticed? Or the opposite, being so fidgety or restless that s/he has been moving around a lot more than usual?

- ☐ 0 = No **[SKIP TO E4i1]**
- ☐ 1 = Yes
- ☐ 9 = **Unknown or unable to assess [SKIP TO E4i1]**

E4h2. SYMPTOM FREQUENCY: Over the last 2 weeks, how often did the patient/resident move or speak so slowly that other people have noticed? Or the opposite, being so fidgety or restless that s/he has been moving around a lot more than usual?

- ☐ 0 = Never or 1 day
- ☐ 1 = 2-6 days (several days)
- ☐ 2 = 7-11 days (half or more of the days)
- ☐ 3 = 12-14 days (nearly every day)
- ☐ 9 = **Unknown or unable to assess**

E4i1. SYMPTOM PRESENCE: Over the last 2 weeks, did the patient/resident state that life isn't worth living, wishes for death, or attempts to harm self?

- ☐ 0 = No **[SKIP TO E4j1]**
- ☐ 1 = Yes
- ☐ 9 = **Unknown or unable to assess [SKIP TO E4j1]**

E4i2. SYMPTOM FREQUENCY: Over the last 2 weeks, how often did the patient/resident state that life isn't worth living, wishes for death, or attempts to harm self?

- ☐ 0 = Never or 1 day
- ☐ 1 = 2-6 days (several days)
- ☐ 2 = 7-11 days (half or more of the days)
- ☐ 3 = 12-14 days (nearly every day)
- ☐ 9 = **Unknown or unable to assess**

E4j1. SYMPTOM PRESENCE: Over the last 2 weeks, was the patient/resident being short-tempered, easily annoyed?

- ☐ 0 = No **[SKIP TO PHQ-9 TOTAL SCORE]**
- ☐ 1 = Yes
- ☐ 9 = **Unknown or unable to assess [SKIP TO PHQ-9 TOTAL SCORE]**

E4j2. SYMPTOM FREQUENCY: Over the last 2 weeks, how often was the patient/resident being short-tempered, easily annoyed?

- ☐ 0 = Never or 1 day
- ☐ 1 = 2-6 days (several days)
- ☐ 2 = 7-11 days (half or more of the days)
- ☐ 3 = 12-14 days (nearly every day)
- ☐ 9 = **Unknown or unable to assess**

PHQ-9-OV TOTAL: Add values from E4a2, E4b2, E4c2, E4d2, E4e2, E4f2, E4g2, E4h2, E4i2, E4j2 →

Testing Objectives

As described in Volume 2, the non-communicative assessments were administered at a single point not directly tied to an admission or discharge to maximize the number of patients/residents eligible for these assessments during the National Beta Test field period. Basic descriptive statistics (e.g., frequencies, means, SDs) are presented for the PHQ-9-OV data elements to characterize the rates of depressed mood for patients/residents in each setting and for the overall sample. We also examined the PHQ-9-OV Total Score by patient/resident characteristics and clinical groups of interest. Feasibility (rates of missingness and time to complete) and interrater reliability (kappa and percent agreement) were examined as well.

Results

Feasibility

Frequencies/Missing

Table 4.1 shows the percentage of responses for each Staff Assessment of Patient/Resident Mood (PHQ-9-OV) data element overall and by setting. The PHQ-9-OV was administered to 501 patients/residents: 32 in HHAs, 98 in IRFs, 155 in LTCHs, and 216 in SNFs. Overall, more than 91 percent of the non-communicative sample was administered the Staff Assessment of Patient/Resident Mood. Among these patients/residents, overall missing data at the data element level ranged from 12.4 percent to 44.3 percent, the majority of which reflected cases in which responses were unknown or unable to be assessed. That said, rates were noticeably lower for data elements asking about sleep interference, moving or speaking slowly, and short temper (12.4 percent to 13.6 percent). At the setting level, missing data rates were generally similar; however, rates were slightly higher in LTCHs compared with other settings (HHAs: 3.1 percent to 37.5 percent, IRFs: 6.1 percent to 38.8 percent, LTCHs: 16.1 percent to 58.7 percent, SNFs: 9.3 percent to 44.4 percent).

Table 4.1. Overall and Setting-Specific Response Frequencies for PHQ-9-OV Data Elements (percent)

Data Element	HHA (*n* = 32)	IRF (*n* = 98)	LTCH (*n* = 155)	SNF (*n* = 216)	Overall (*n* = 501)
Symptom presence and frequency: little interest or pleasure (e4a)					
No	71	60	63	66	65
0–1 day	4	5	2	4	4
2–6 days	17	7	5	4	6
7–11 days (half or more)	0	7	5	1	3
12–14 days (nearly all)	8	22	27	25	23
Symptom presence and frequency: feeling down, depressed, hopeless (e4b)					
No	69	62	61	79	70
0–1 day	4	4	3	1	2
2–6 days	15	13	5	5	8
7–11 days (half or more)	4	8	5	3	5
12–14 days (nearly all)	8	14	25	13	16

Data Element	HHA (*n* = 32)	IRF (*n* = 98)	LTCH (*n* = 155)	SNF (*n* = 216)	Overall (*n* = 501)
Symptom presence and frequency: too little/too much sleep (e4c)					
No	64	55	52	75	63
0–1 day	4	8	3	2	3
2–6 days	11	17	9	4	8
7–11 days (half or more)	7	5	6	3	5
12–14 days (nearly all)	14	16	29	17	20
Symptom presence and frequency: tired/no energy (e4d)					
No	44	50	41	64	54
0–1 day	4	4	2	1	2
2–6 days	19	18	8	5	10
7–11 days (half or more)	11	7	10	5	7
12–14 days (nearly all)	22	21	40	25	27
Symptom presence and frequency: poor appetite or overeating (e4e)					
No	77	59	81	68	71
0–1 day	0	1	1	2	1
2–6 days	10	18	4	6	8
7–11 days (half or more)	6	4	1	5	4
12–14 days (nearly all)	6	18	13	19	16
Symptom presence and frequency: feel bad about self (e4f)					
No	90	82	99	96	93
0–1 day	0	2	0	1	1
2–6 days	5	5	0	1	2
7–11 days (half or more)	5	0	0	2	1
12–14 days (nearly all)	0	11	1	1	3
Symptom presence and frequency: trouble concentrating (e4g)					
No	59	34	57	61	53
0–1 day	0	5	0	0	1
2–6 days	9	14	1	2	6
7–11 days (half or more)	5	8	9	2	6
12–14 days (nearly all)	27	39	33	34	34

Data Element	HHA (*n* = 32)	IRF (*n* = 98)	LTCH (*n* = 155)	SNF (*n* = 216)	Overall (*n* = 501)
Symptom presence and frequency: moving or speaking slowly (e4h)					
No	69	62	62	81	71
0–1 day	0	3	2	1	1
2–6 days	10	9	8	5	7
7–11 days (half or more)	7	4	5	3	4
12–14 days (nearly all)	14	23	22	10	16
Symptom presence and frequency: suicidal thoughts (e4i)					
No	100	97	97	98	98
0–1 day	0	0	1	0	0
2–6 days	0	3	1	0	1
7–11 days (half or more)	0	0	0	1	0
12–14 days (nearly all)	0	0	1	1	1
Symptom presence and frequency: short-tempered (e4j)					
No	74	75	81	74	76
0–1 day	3	2	0	2	2
2–6 days	3	10	3	7	6
7–11 days (half or more)	3	4	3	4	4
12–14 days (nearly all)	16	9	13	13	12
PHQ-9-OV Total Score					
Mean (SD)	5.5 (4.9)	7.1 (6.9)	7.8 (7.6)	5.8 (7.0)	6.6 (7.1)
Depression categorization (PHQ-9-OV)					
No depression	53	60	66	64	63
Minor depression	33	15	25	23	22
Major depression	13	25	10	13	15

The Staff Assessment of Patient/Resident Mood results show that 63 percent of patients/residents overall were categorized as having no depression, 22 percent as having risk for minor depression, and 15 percent as having risk for major depression based on their total severity scores. As expected, setting type was associated with depressed mood $\left(\chi^2_{(6)} = 13.50, p < 0.05\right)$, such that a higher percentage of patients/residents who fall into the category of major depression were found in the IRF setting (25 percent), compared with 13 percent in HHAs and SNFs and 10 percent in LTCHs. Despite the fact that 66 percent of patients/residents at LTCHs are in the no depression category (the highest compared with other settings) and only 10 percent are in the major depression category (the lowest compared with other settings), LTCH patients as a group had the highest average PHQ-9-OV score (M = 7.8, SD = 7.6) compared with those in HHAs (M

= 5.5, SD = 4.9), IRFs (M = 7.1, SD = 6.9), and SNFs (M = 5.8, SD = 7.0), indicating that scores are quite high among those who are in the major depression category. It should be noted, however, that overall PHQ-9-OV scores (and thus depression categorization) tended to be positively skewed, such that nearly two-thirds of patients/residents had relatively low PHQ-9-OV scores (no depression), fewer with moderate scores (minor depression), and even fewer with high scores (major depression). This trend toward lower PHQ-9-OV scores is observed in the relatively low average PHQ-9-OV score with a slightly larger SD, which characterizes the overall spread. Moreover, this general distributional pattern was observed in all settings except IRFs, where, as noted previously, the second-largest group consisted of patients/residents with major depression.

Exploratory Comparisons with Known Groups

In other volumes of this report, we present associations between patient/resident performance on the candidate SPADEs and other known patient/resident characteristics. In those cases, we formed expectations about the associations based on the available research literature on populations who are similar to the Beta sample (e.g., patients/residents receiving PAC, older adults, nursing home residents). Observing expected or logical associations contributed to evidence that the data elements are valid—that is, that they assess the construct that they are intended to capture. However, our ability to generate hypotheses for the non-communicative sample was limited, primarily because of the lack of research on equivalent populations (i.e., patients/residents receiving PAC services who are unable to communicate by any means). In addition, the smaller sample of non-communicative patients/residents limits the power of these analyses to detect differences between rates of less common conditions or characteristics, such as sepsis or certain discharge disposition categories. For these reasons, we consider the comparisons with known groups described below to be exploratory. We present them as part of this volume for completeness but not as evidence for or against the construct validity of the candidate SPADEs.

Table 4.2 shows means and SDs to the PHQ-9-OV Total Scores for the overall non-communicative 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 or older), 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 the 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).

Table 4.2. Overall Mean (SD) PHQ-9-OV Total Score by Patient/Resident Characteristics and Clinical Groups

Patient/Resident Characteristics and Clinical Groups	Mean PHQ-9-OV Total Score (SD)
Gender (*n* = 416)	
Male (*n* = 182)	6.5 (6.8)
Female (*n* = 234)	6.6 (7.3)
Age (*n* = 415)	
18–44 (*n* = 20)	6.5 (7.3)
45–64 (*n* = 79)	7.4 (7.4)
65–74 (*n* = 87)	6.7 (7.2)
75–89 (*n* = 163)	6.3 (6.8)
90 or older (*n* = 66)	6.0 (7.3)
Length of stay (*n* = 214)	Pearson *r* = 0.02
Disposition at discharge (*n* = 408)	
Home (*n* = 54)	5.1 (5.6)
Hospital (*n* = 35)	7.5 (6.4)
Hospice (*n* = 13)	12.0 (9.0)
HHA (*n* = 27)	6.3 (5.2)
IRF (*n* = 12)	7.9 (7.7)
LTCH (*n* = 0)	N/A
SNF (*n* = 117)	6.9 (7.5)
Other (*n* = 150)	6.4 (7.3)
Clinical conditions (*n* = 252)	
Sepsis	
Yes (*n* = 30)	8.5 (6.8)
No (*n* = 222)	6.3 (7.0)
Heart failure	
Yes (*n* = 16)	3.8 (5.5)
No (*n* = 236)	6.7 (7.1)
Stroke	
Yes (*n* = 63)	6.7 (7.3)
No (*n* = 189)	6.5 (6.9)

Patient/Resident Characteristics and Clinical Groups	Mean PHQ-9-OV Total Score (SD)
Hygiene—Toileting (*n* = 161)	
Independent (*n* = 2)	7.5 (10.6)
Setup or clean-up assistance (*n* = 1)	10.0 (N/A)
Supervision or touching assistance (*n* = 5)	3.8 (3.7)
Partial/moderate assistance (*n* = 10)	7.3 (8.7)
Substantial/maximal assistance (*n* = 21)	5.9 (5.4)
Dependent (*n* = 122)	7.8 (7.9)
Mobility—Transfer from lying to sitting (*n* = 153)	
Independent (*n* = 4)	6.3 (7.5)
Setup or clean-up assistance (*n* = 1)	0.0 (N/A)
Supervision or touching assistance (*n* = 14)	5.9 (6.3)
Partial/moderate assistance (*n* = 24)	6.2 (6.6)
Substantial/maximal assistance (*n* = 32)	8.9 (8.7)
Dependent (*n* = 78)	7.8 (7.2)

NOTE: N/A = not applicable.

Because of the heterogeneous nature of the non-communicative population and a lack of research on equivalent populations, we did not form hypotheses or expectations about the associations among patient/resident characteristics. In fact, across all patient/resident characteristics and clinical conditions, there were no significant associations with patients'/residents' PHQ-9-OV Total Scores. As a reminder, PHQ-9-OV Total Scores were positively skewed, such that nearly two-thirds of all patients/residents had low PHQ-9-OV Total Scores (no depression). It is not surprising that no significant associations emerged given this substantial clustering of low PHQ-9-OV Total Scores.

Time to Complete

Table 4.3 shows the average time to complete the PHQ-9-OV data elements overall and by setting. On average, the time to complete was 3.5 minutes (SD = 1.7) and ranged from 3.3 minutes (SD = 1.5) in SNFs to 4.5 minutes in HHAs (SD = 2.4). There were differences in time to complete at the setting level ($F_{(3,471)} = 4.42$, $p < 0.01$), indicating that the PHQ-9-OV tended to take significantly more time in HHAs than in SNFs ($t_{(471)} = 3.46$, $p < 0.001$), IRFs ($t_{(471)} = 5.22$, $p < 0.05$) and LTCHs ($t_{(471)} = 8.51$, $p < 0.01$).

Table 4.3. Time to Complete for PHQ-9-OV Data Elements (minutes)

Time to Complete	HHA (*n* = 27)	IRF (*n* = 99)	LTCH (*n* = 138)	SNF (*n* = 211)	Overall (*n* = 475)
Mean (SD)	4.5 (2.4)	3.7 (1.9)	3.5 (1.8)	3.3 (1.5)	3.5 (1.7)

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.9–A.12 in the appendix). The PHQ-9-OV data elements took significantly less time to complete in the Midwest region (3.1 minutes [SD = 1.5]) than in the South (3.6 minutes [SD = 1.7], Cohen's $d = 0.31$) or the West (4.0 minutes [SD = 1.8], Cohen's $d = 0.54$). These data elements also took significantly less time to complete in the Northeast region (3.4 minutes [SD = 1.9]) than in the West (Cohen's $d = 0.32$). There were no other significant differences in time to complete the PHQ-9-OV data elements in these sensitivity analyses.

Interrater Reliability

Table 4.4 shows the kappa interrater reliability coefficients for the Staff Assessment of Patient/Resident Mood (PHQ-9-OV) data elements overall and by setting. Kappa coefficients were computed on 487 patients/residents. Overall kappa coefficients were excellent, ranging from 0.92 to 0.98. At the setting level, kappa coefficients were good to excellent and ranged from 0.84 to 1.00 in HHAs, 0.91 to 1.00 in IRFs, 0.89 to 0.99 in LTCHs, and 0.95 to 1.00 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.13–A.16 in the appendix). No noteworthy differences were found for interrater reliability of the Staff Assessment of Patient/Resident Mood (PHQ-9-OV) data elements in these sensitivity analyses.

Table 4.4. Interrater Reliability Kappa or Weighted Kappa for PHQ-9-OV Data Elements

Data Element	HHA (*n* = 32)	IRF (*n* = 92)	LTCH (*n* = 153)	SNF (*n* = 210)	Overall (*n* = 487)
Symptom presence and frequency: little interest or pleasure (e4a)	1.00	0.92	0.99	0.97	0.97
Symptom presence and frequency: feeling down, depressed, hopeless (e4b)	1.00	0.95	0.99	0.99	0.98
Symptom presence and frequency: too little/too much sleep (e4c)	1.00	0.97	0.97	1.00	0.98
Symptom presence and frequency: tired/no energy (e4d)	1.00	0.99	0.96	0.97	0.98
Symptom presence and frequency: poor appetite or overeating (e4e)	0.94	0.92	0.90	0.96	0.95
Symptom presence and frequency: feel bad about self (e4f)	1.00	0.98	—	—	—
Symptom presence and frequency: trouble concentrating (e4g)	0.91	0.91	0.90	0.97	0.94
Symptom presence and frequency: moving or speaking slowly (e4h)	0.84	0.94	0.89	0.95	0.92
Symptom presence and frequency: suicidal thoughts (e4i)	—	—	—	—	—
Symptom presence and frequency: short-tempered (e4j)	1.00	1.00	0.92	0.99	0.97
Sum of all symptom frequencies (PHQ-9-OV)[a]	1.00	0.98	0.92	0.98	0.97

NOTES: Interrater reliability 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, and 0.81–1.00 is excellent/almost perfect.
[a] As classified into the three categories shown in Table 4.1 (no depression, minor depression, major depression).

Table 4.5 shows percent agreement for the Staff Assessment of Patient/Resident Mood (PHQ-9-OV) data elements overall and by setting. Overall percent agreement was high for all data elements, ranging from 96 percent to 99 percent, with minimal setting differences. At the setting level, percent agreement was high for all data elements and ranged from 93 percent to 100 percent in HHAs, IRFs, and LTCHs, and from 97 percent to 100 percent in SNFs.

Table 4.5. Interrater Reliability—Percent Agreement for PHQ-9-OV Data Elements

Data Element	HHA (*n* = 32)	IRF (*n* = 92)	LTCH (*n* = 153)	SNF (*n* = 210)	Overall (*n* = 487)
Symptom presence and frequency: little interest or pleasure (e4a)	100	93	98	97	97
Symptom presence and frequency: feeling down, depressed, hopeless (e4b)	100	96	99	99	98
Symptom presence and frequency: too little/too much sleep (e4c)	100	97	97	99	98
Symptom presence and frequency: tired/no energy (e4d)	100	99	96	97	98
Symptom presence and frequency: poor appetite or overeating (e4e)	97	95	96	97	97
Symptom presence and frequency: feel bad about self (e4f)	100	98	100	99	99
Symptom presence and frequency: trouble concentrating (e4g)	95	94	93	98	96
Symptom presence and frequency: moving or speaking slowly (e4h)	93	96	94	97	96
Symptom presence and frequency: suicidal thoughts (e4i)	100	100	98	100	99
Symptom presence and frequency: short-tempered (e4j)	100	100	97	98	99
Sum of all symptom frequencies (PHQ-9-OV)[a]	100	99	95	98	98

[a] As classified into the three categories shown in Table 4.1 (no depression, minor depression, major depression).

Assessor Feedback

Although facility staff and research nurses considered standardized assessment of signs and symptoms of depression to be important, they did not provide feedback specific to the observational assessment in the focus groups. According to the assessor survey, facility staff and research nurses rated the Staff Assessment of Patient/Resident Mood as somewhat to moderately clinically useful and reported a higher data collection burden than many of the other data elements.

Summary

Results for the Staff Assessment of Patient/Resident Mood indicate moderate overall support for cross-setting standardization. Assessors considered the Staff Assessment of Patient/Resident Mood data elements to be only somewhat clinically useful, with a high data collection burden compared with the other data elements. There were no significant associations between the total score of the PHQ-9-OV and patient/resident characteristics; this is likely due to the low rates of depressive symptoms observed among this population.

In terms of feasibility, the rates of missing data were high for many of the data elements in this assessment. These rates are due to the staff observer's inability to determine an accurate answer based on his or her observation of the patient/resident, which is an inherent challenge with all observational assessments with non-communicative patients/residents. Notably, the three data elements with relatively lower missing rates assessed symptoms that are most readily observed (sleep, movement, and temper). Time to complete was 3.5 minutes on average and took significantly longer in the HHA setting relative to other settings. Additionally, the data elements were completed more quickly in the Midwest and Northeast relative to the West and South regions. These effects exceeded our cutoff effect size value of 0.2, raising some questions about the generalizability of the time-to-complete estimate. Psychometric performance of the data elements was excellent: Kappas for the Staff Assessment of Patient/Resident Mood were excellent and percent agreement was high for all data elements. These combined results show moderately low feasibility (but excellent interrater reliability) and somewhat to moderate clinical utility (but high data collection burden) for the Staff Assessment of Patient/Resident Mood as a candidate data element for standardization across PAC settings.

5. Observational Assessment of Pain or Distress

Data Element Description

The data elements that constitute the Observational Assessment of Pain or Distress collect staff observations of patients'/residents' expressed behavioral indicators of potential pain or distress. These data elements were designed to be administered to all patients/residents who are unable to communicate (i.e., cannot reliably make themselves understood via verbal communication, written communication, a communication board, eye blinks, etc.).

The Observational Assessment of Pain or Distress data elements are completed through interviews with staff, family members, caregivers, a review of the medical record, and observation of the patient/resident during care activities. A similar set of data elements is currently included in the MDS. These data elements are shown in Figure 5.1.

Figure 5.1. Observational Assessment of Pain or Distress Data Elements

D7. Observational Assessment of Pain or Distress

FOR ALL PATIENTS/RESIDENTS WHO ARE UNABLE TO PARTICIPATE IN THE PAIN INTERVIEW, PLEASE NOTE WHETHER ANY OF THE FOLLOWING BEHAVIORS WERE OBSERVED.

PATIENTS/RESIDENTS SHOULD BE OBSERVED TWICE DAILY (MORNING AND EVENING) DURING CARE ACTIVITIES (I.E., DURING TRANSFER PROCEDURES, REPOSITIONING, BATHING, TOILETING, WOUND CARE/DRESSING CHANGES, RANGE OF MOTION, AMBULATING, OR OTHER EXERCISES, ETC.), WHEN BEHAVIORAL SIGNS OF POTENTIAL PAIN OR DISTRESS ARE MOST LIKELY TO BE EXPRESSED, OVER THE COURSE OF 3 CONSECUTIVE DAYS.

CHECK ALL THAT APPLY

- ☐ a = Non-verbal sounds (e.g., crying, whining, gasping, moaning, or groaning)
- ☐ b = Vocal complaints of pain (e.g., "that hurts, ouch, stop")
- ☐ c = Facial expressions (e.g., grimaces, winces, wrinkled forehead, furrowed brow, clenched teeth or jaw, rapid eye blinking, tightly closed eyes)
- ☐ d = Body movements or postures (e.g., bracing, guarding, rubbing or massaging a body part/area; clutching or holding a body part during movement; rigid, tense body posture; withdrawing an extremity to an external stimulus; fidgeting; increased pacing, rocking; restricted movement; gait or mobility changes)
- ☐ z = None of these signs observed or documented. **[SKIP TO DNC-TIME]**

D8. For patients/residents who demonstrated <u>any</u> indicators of potential pain or distress listed in D7: Observational Assessment of Pain or Distress, identify the frequency with which patient/resident complains or shows evidence of potential pain or distress <u>over the past 3 days</u>.

- ☐ 1 = Indicators of potential pain or distress observed less than daily
- ☐ 2 = Indicators of potential pain or distress observed daily (at least once per day on each day of the assessment window)
- ☐ 3 = Indicators of potential pain or distress observed more than daily (multiple times per day on each day of the assessment window)
- ☐ 9 = **Unknown or unable to assess**

D9. For patients/residents who demonstrated any indicators of potential pain or distress listed in D7: Observational Assessment of Pain or Distress, is there any evidence that these indicators resolved or diminished in response to pain medications or treatments <u>over the past 3 days</u>?

- ☐ 0 = No
- ☐ 1 = Yes
- ☐ 8 = **Not applicable** – patient/resident has not received pain medications or treatments within the past 3 days
- ☐ 9 = **Unknown or unable to assess**

Testing Objectives

As described in Volume 2, the non-communicative assessments were administered at a single point not directly tied to an admission or discharge to maximize the number of patients/residents eligible for these assessments during the National Beta Test field period. Basic descriptive statistics (e.g., frequencies, means, SDs) are presented for the Observational Assessment of Pain

or Distress data elements to characterize the rates of pain for patients/residents in each setting and for the overall sample. We also examined the rates of any signs of pain observed (i.e., checking boxes a, b, c, or d in data element d7 in Figure 5.1) by patient/resident characteristics and clinical groups of interest. Feasibility (rates of missingness and time to complete) and interrater reliability (kappa and percent agreement) were examined as well.

Results

Feasibility

Frequencies/Missing

Table 5.1 shows the percentage of responses for the Observational Assessment of Pain or Distress data elements overall and by setting. The Observational Assessment of Pain or Distress was administered to 545 patients/residents: 32 in HHAs, 107 in IRFs, 183 in LTCHs, and 223 in SNFs. Overall, more than 99 percent of the non-communicative sample was administered the Observational Assessment of Pain or Distress. Among these patients/residents, overall missing data at the data element level ranged from 0.0 percent to 4.4 percent. At the setting level, missing data rates were generally similar (HHAs: 0.0 percent to 6.3 percent, IRFs: 0.0 percent to 7.5 percent, LTCHs: 0.0 percent to 3.8 percent, SNFs: 0.0 percent to 4.5 percent).

The Observational Assessment of Pain or Distress results show that the most frequent sign of pain or distress overall was facial expressions (37 percent); the other symptoms were observed among 23 percent to 27 percent of patients/residents. However, it was most common for none of the pain or distress signs (i.e., vocal complaints, facial expression, or body movements or postures) to be observed or documented (44 percent of patients/residents). Setting type was associated with pain or distress $\left(\chi^2_{(9)} = 30.16, p < 0.001\right)$, such that a higher percentage of patients/residents who showed none of these pain or distress signs were found in IRF and SNF settings (50 percent of the total sample in each setting, compared with 31 percent and 36 percent of patients/residents in HHAs and LTCHs, respectively).

All of the data elements in d7 (a, b, c, and d) tended to be observed at higher rates in HHAs compared with other settings, and the data elements focusing on facial expressions and body movements (d7c, d7d) were observed at higher rates for LTCH patients and HHA patients relative to those in IRF and SNF settings. Indicators of pain having resolved or diminished in response to pain medications or treatments over the past three days was significantly higher in HHA settings (all 22 of the HHA patients showing signs of pain according to d7 showed some indication of pain resolution because of treatment) than in the other settings (81 percent to 83 percent).

Table 5.1. Overall and Setting-Specific Response Frequencies for the Observational Assessment of Pain or Distress Data Elements (percent)

Data Element	HHA (*n* = 32)	IRF (*n* = 107)	LTCH (*n* = 183)	SNF (*n* = 223)	Overall (*n* = 545)
Observed pain or distress: nonverbal sounds (d7a)					
Yes	34	24	21	23	23
Observed pain or distress: vocal complaints (d7b)					
Yes	41	24	19	22	23
Observed pain or distress: facial expressions (d7c)					
Yes	44	35	48	30	37
Observed pain or distress: body movements or postures (d7d)					
Yes	50	22	35	19	27
Observed pain or distress: none observed or documented (d7)					
Yes	31	50	36	50	44
Frequency of patient complaints or evidence of pain in past three days (d8)					
Less than daily	30	55	32	52	43
Daily	25	22	24	19	22
More than daily	45	24	43	29	35
Did indicators of pain resolve/diminish with pain medications or treatment (d9)					
Yes	100	83	83	81	83

Exploratory Comparisons with Known Groups

In other volumes of this report, we present associations between patient/resident performance on the candidate SPADEs and other known patient/resident characteristics. In those cases, we formed expectations about the associations based on the available research literature on populations who are similar to the Beta sample (e.g., patients/residents receiving PAC, older adults, nursing home residents). Observing expected or logical associations contributed to evidence that the data elements are valid—that is, that they assess the construct that they are intended to capture. However, our ability to generate hypotheses for the non-communicative sample was limited, primarily because of the lack of research on equivalent populations (i.e., patients/residents receiving PAC services who are unable to communicate by any means). In addition, the smaller sample of non-communicative patients/residents limits the power of these analyses to detect differences between rates of less common conditions or characteristics, such as sepsis or certain discharge disposition categories. For these reasons, we consider the comparisons with known groups described below to be exploratory. We present them as part of this volume

for completeness but not as evidence for or against the construct validity of the candidate SPADEs.

Table 5.2 shows the frequencies for any signs of pain observed for the overall non-communicative 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 or older), 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).

Table 5.2. Overall Frequencies for Any Signs of Pain Observed by Patient/Resident Characteristics and Clinical Groups (percent)

Patient/Resident Characteristics and Clinical Groups	Any Signs of Pain Observed (Yes)
Gender (*n* = 536)	
Male (*n* = 245)	48.6
Female (*n* = 291)	40.6
Age (*n* = 534)	
18–44 (*n* = 33)	57.6
45–64 (*n* = 100)	37.0
65–74 (*n* = 108)	45.4
75–89 (*n* = 209)	44.0
90 or older (*n* = 84)	47.6
Length of stay (*n* = 278; mean, SD)	Yes: 24.5 (13.4) No: 25.6 (13.3)
Disposition at discharge (*n* = 518)	
Home (*n* = 59)	54.2
Hospital (*n* = 52)	46.2
Hospice (*n* = 20)	25.0
SNF (*n* = 149)	40.9
IRF (*n* = 17)	29.4
HHA (*n* = 32)	37.5
LTCH (*n* = 0)	N/A
Other (*n* = 189)	48.7

Patient/Resident Characteristics and Clinical Groups	Any Signs of Pain Observed (Yes)
Clinical conditions (*n* = 339)	
Sepsis	
Yes (*n* = 54)	33.3
No (*n* = 285)	44.9
Heart failure	
Yes (*n* = 24)	58.3
No (*n* = 315)	41.9
Stroke	
Yes (*n* = 86)	44.2
No (*n* = 253)	42.7
Hygiene—Toileting (*n* = 228)	
Independent (*n* = 2)	50.0
Setup or clean-up assistance (*n* = 1)	0.0
Supervision or touching assistance (*n* = 5)	100.0
Partial/moderate assistance (*n* = 11)	63.6
Substantial/maximal assistance (*n* = 28)	39.3
Dependent (*n* = 181)	38.1
Mobility—Transfer from lying to sitting (*n* = 200)	
Independent (*n* = 4)	75.0
Setup or clean-up assistance (*n* = 1)	0.0
Supervision or touching assistance (*n* = 16)	62.5
Partial/moderate assistance (*n* = 28)	57.1
Substantial/maximal assistance (*n* = 39)	38.5
Dependent (*n* = 112)	33.9

NOTE: N/A = not applicable.

Because of the heterogeneous nature of the non-communicative population and a lack of research on equivalent populations, we did not form hypotheses or expectations about the associations among patient/resident characteristics. Across all patient/resident characteristics and clinical conditions, there were no significant associations with patient/resident observed pain.

Time to Complete

Table 5.3 shows the average time to complete the Observational Assessment of Pain or Distress data elements overall and by setting. On average, the time to complete was 2.4 minutes (SD = 1.7) and ranged from 2.3 minutes (SD = 1.6) in SNFs to 2.5 minutes in LTCHs (SD = 1.6). There were no differences in time to complete according to setting.

Table 5.3. Time to Complete the Observational Assessment of Pain or Distress Data Elements (minutes)

Time to Complete	HHA (*n* = 32)	IRF (*n* = 103)	LTCH (*n* = 147)	SNF (*n* = 218)	Overall (*n* = 500)
Mean (SD)	2.4 (1.7)	2.4 (1.9)	2.5 (1.6)	2.3 (1.6)	2.4 (1.7)

Time to complete was also evaluated using the data 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 A.17–A.20 in the appendix). The Observational Assessment of Pain or Distress data elements took significantly less time to complete in the Midwest region (2.1 minutes [SD = 1.5]) than in the South (2.5 minutes [SD = 1.6], Cohen's d = 0.26), the West (2.5 minutes [SD = 1.7], Cohen's d = 0.25) or the Northeast (2.6 minutes [SD = 1.9], Cohen's d = 0.29). There were no other significant differences in time to complete the Observational Assessment of Pain or Distress data elements in these sensitivity analyses.

Interrater Reliability

Table 5.4 shows the kappa interrater reliability coefficients for the Observational Assessment of Pain or Distress data elements overall and by setting. Kappa coefficients were computed on 543 patients/residents. Overall kappa coefficients were excellent, ranging from 0.81 to 0.90. At the setting level, kappa coefficients were good to excellent and ranged from 0.80 to 1.00 in HHAs, 0.65 to 0.80 in IRFs, 0.80 to 0.91 in LTCHs, and 0.85 to 0.95 in SNFs.

Table 5.4. Interrater Reliability Kappa or Weighted Kappa for Observational Assessment of Pain or Distress Data Elements

Data Element	HHA (*n* = 32)	IRF (*n* = 106)	LTCH (*n* = 183)	SNF (*n* = 222)	Overall (*n* = 543)
Observed pain or distress: nonverbal sounds (d7)	0.86	0.65	0.80	0.87	0.81
Observed pain or distress: vocal complaints (d7)	0.94	0.73	0.87	0.90	0.86
Observed pain or distress: facial expressions (d7)	0.80	0.76	0.82	0.85	0.83
Observed pain or distress: body movements or postures (d7)	0.88	0.80	0.80	0.87	0.83
Observed pain or distress: none observed or documented (d7)	1.00	0.75	0.91	0.92	0.89
Frequency of patient complaints or evidence of pain in past three days (d8)	0.89	0.78	0.90	0.85	0.87
Did indicators of pain resolve/diminish with pain medications or treatment (d9)	—	—	0.82	0.95	0.90

NOTES: Interrater reliability 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, and 0.81–1.00 is excellent/almost perfect.

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). The Northeast region had a substantially lower kappa for the nonverbal sounds data element (0.40) relative to the other regions (0.80 to 0.88). No other noteworthy differences were found for interrater reliability of the Observational Assessment of Pain or Distress data elements in these sensitivity analyses.

Table 5.5 shows percent agreement for the Observational Assessment of Pain or Distress data elements overall and by setting. Overall percent agreement was high for all data elements, ranging from 89 percent to 98 percent, with minimal setting differences. At the setting level, percent agreement was high for all data elements and ranged from 90 percent to 100 percent in HHAs, 83 percent to 100 percent in IRFs, 91 percent to 96 percent in LTCHs, and 88 percent to 99 percent in SNFs.

Table 5.5. Interrater Reliability—Percent Agreement for Observational Assessment of Pain or Distress Data Elements

Data Element	HHA (*n* = 32)	IRF (*n* = 106)	LTCH (*n* = 183)	SNF (*n* = 222)	Overall (*n* = 543)
Observed pain or distress: nonverbal sounds (d7)	94	88	93	95	93
Observed pain or distress: vocal complaints (d7)	97	91	96	96	95
Observed pain or distress: facial expressions (d7)	91	90	91	94	92
Observed pain or distress: body movements or postures (d7)	94	92	91	96	93
Observed pain or distress: none observed or documented (d7)	100	88	96	96	94
Frequency of patient complaints or evidence of pain in past three days (d8)	90	83	91	88	89
Did indicators of pain resolve/diminish with pain medications or treatment (d9)	100	100	96	99	98

Assessor Feedback

Although facility staff and research nurses considered pain to be a clinically useful measurement across PAC settings, they did not provide feedback specific to the observational assessment in the focus groups. According to the assessor survey, facility staff and research nurses found the observational assessment of pain or distress to be somewhat to moderately clinically useful and thought that it was only slightly difficult to collect information for this data element.

Summary

Results for the Observational Assessment of Pain or Distress indicate moderate overall support for cross-setting standardization. Assessors considered Observational Assessment of Pain or Distress data elements to be somewhat clinically useful and to have a moderate data collection burden. Although there were no significant associations between any signs of pain observed and patient/resident characteristics, this finding is likely due to the smaller sample size of the non-communicative sample in the National Beta Test.

In terms of feasibility, the rates of missing data were reasonable for the data elements in the pain assessment, in contrast to the other two observational assessments for non-communicative patients/residents. This difference is likely due to the readily observed behaviors included in the pain assessment. Time to complete was under 2.5 minutes on average with no setting differences. However, the data elements were completed more quickly in the Midwest relative to the other three regions. These effects were small but exceeded our cutoff effect size value of 0.2, raising some questions about the generalizability of the time-to-complete estimate. In terms of psychometric performance, kappas for the Observational Assessment of Pain or Distress were excellent overall, and percent agreement was high for all data elements. These combined results show moderate feasibility, excellent interrater reliability, somewhat to moderate clinical utility, and moderate data collection burden for the Observational Assessment of Pain or Distress as a candidate data element for standardization across PAC settings.

6. Conclusion

The National Beta Test evaluated several candidate-standardized patient assessment data elements developed specifically for patients/residents who are unable to communicate. These data element sets included the Staff Assessment of Mental Status to assess cognitive function, the Staff Assessment of Patient/Resident Mood to assess mental status, and the Observational Assessment of Pain or Distress to assess pain. The rates of cognitive impairment and depressive symptoms reported in this volume for the non-communicative patients/residents were slightly higher than would be observed in the general PAC patient/resident population (i.e., one that includes both communicative and non-communicative patients/residents) because of the clinical differences in these patients/residents.

The general performance of these three data element sets is summarized for the full non-communicative sample (combined across settings) in Table 6.1. As can be seen in Table 6.1, the data elements performed reasonably well, showing feasibility, acceptable reliability, and support from assessors. However, there are some differences in performance among the three that are worthy of consideration. Specifically, in terms of feasibility, rates of missing data were high for many of the data elements in the Staff Assessment of Mental Status and Staff Assessment of Patient/Resident Mood but were reasonable for the Observational Assessment of Pain or Distress. Of the three observational assessments, the Staff Assessment of Patient/Resident Mood took the longest to complete (mean = 3.5 minutes, SD = 1.7). The Staff Assessment of Mental Status and the Observational Assessment of Pain or Distress took an average of 2.6 minutes (SD = 1.6) and 2.4 minutes (SD = 1.7) to complete, respectively. Furthermore, the time to complete was longer in HHAs than in other settings, perhaps because of HHA assessors' limited familiarity with the patients at the time of the assessment (i.e., based on only one or two visits).

Interrater reliability was good to excellent for all data elements, although some data elements performed better than others. The variability in data element performance seemed to align with the content of the data elements, such that those symptoms that were more readily observed were documented more reliably. Overall kappas for the Staff Assessment of Patient/Resident Mood and Observational Assessment of Pain or Distress were excellent (ranging from 0.92 to 0.98 and 0.81 to 0.90, respectively). Kappas for the data elements in the Staff Assessment of Mental Status were also excellent, except for one data element ("short-term memory OK") with a kappa of 0.74 (good). These data elements also showed excellent interrater reliability based on percent agreement.

Table 6.1. Observational Assessments of Cognitive Function, Mental Status, and Pain: Summary of Data Element Performance in National Beta Test (Combined Sample)

Data Element	Mean (SD) Time to Complete (minutes)	Interrater Reliability		Assessor Feedback
		Kappa	Percent Agreement	
Staff Assessment of Mental Status	2.6 (1.6)	0.74–0.94	93–98%	Somewhat to moderate clinical utility, moderate burden
PHQ-9-OV	3.5 (1.7)	0.92–0.98	96–99%	Somewhat to moderate clinical utility, high burden
Observational Assessment of Pain or Distress	2.4 (1.7)	0.81–0.90	89–98%	Moderate to high clinical utility, moderate burden

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, and 0.81–1.00 is excellent/almost perfect.

Assessor feedback for these three data elements was limited, and available feedback was somewhat mixed for the three observational assessments. The data elements were all deemed to be at least somewhat to moderately clinically useful but with a moderate to high burden by the clinical assessors in this study. Although the assessors did not provide feedback on the specific data elements in focus groups, focus group participants noted that it is difficult for staff to report on patients/residents for the observational assessments if the staff are assigned to the patient/resident for only some portion of the assessment period.

Appendix. Supplementary Tables

Supplementary Tables for Staff Assessment of Mental Status

Table A.1. Time to Complete the Staff Assessment of Mental Status Data Elements by Urbanicity (minutes)

Time to Complete	Urban (*n* = 471)	Nonurban (*n* = 18)	Overall (*n* = 489)
Mean (SD)	2.6 (1.6)	2.8 (2.2)	2.6 (1.6)

Table A.2. Time to Complete the Staff Assessment of Mental Status Data Elements by Region (minutes)

Time to Complete	Northeast (*n* = 85)	South (*n* = 158)	Midwest (*n* = 133)	West (*n* = 113)	Overall (*n* = 489)
Mean (SD)	2.5 (1.6)	2.7 (1.7)	2.3 (1.5)	2.7 (1.5)	2.6 (1.6)

Table A.3. Time to Complete the Staff Assessment of Mental Status Data Elements by Facility Ownership (minutes)

Time to Complete	For-Profit (*n* = 346)	Nonprofit (*n* = 141)	Overall (*n* = 489)
Mean (SD)	2.5 (1.5)	2.8 (1.7)	2.6 (1.6)

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

Table A.4. Time to Complete the Staff Assessment of Mental Status Data Elements by Facility Size (minutes)

Time to Complete	Below Setting-Type Median (*n* = 258)	Above Setting-Type Median (*n* = 230)	Overall (*n* = 489)
Mean (SD)	2.5 (1.5)	2.6 (1.7)	2.6 (1.6)

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

Table A.5. Interrater Reliability Kappa or Weighted Kappa for Staff Assessment of Mental Status Data Elements by Urbanicity

Data Element	Urban (*n* = 486)	Nonurban (*n* = 19)
Short-term memory OK (b3a)	0.72	1.00
Long-term memory OK (b3b)	0.81	1.00
Is the patient normally able to recall: current season (b3c1)	0.90	1.00
Is the patient normally able to recall: location of own room (b3c2)	0.93	1.00
Is the patient normally able to recall: staff names and faces (b3c3)	0.89	0.90
Is the patient normally able to recall: that he or she is in a care facility (b3c4)	0.89	1.00
Ability to make decisions regarding everyday tasks (b3d)	0.89	1.00

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, and 0.81–1.00 is excellent/almost perfect.

Table A.6. Interrater Reliability Kappa or Weighted Kappa for Staff Assessment of Mental Status Data Elements by Region

Data Element	Northeast (*n* = 69)	South (*n* = 164)	Midwest (*n* = 129)	West (*n* = 143)
Short-term memory OK (b3a)	0.39	0.83	0.79	0.72
Long-term memory OK (b3b)	0.93	0.73	0.89	0.83
Is the patient normally able to recall: current season (b3c1)	0.55	—	1.00	0.94
Is the patient normally able to recall: location of own room (b3c2)	0.81	0.95	1.00	0.91
Is the patient normally able to recall: staff names and faces (b3c3)	0.70	0.85	0.98	0.92
Is the patient normally able to recall: that he or she is in a care facility (b3c4)	0.66	0.89	0.98	0.91
Ability to make decisions regarding everyday tasks (b3d)	0.64	0.92	1.00	0.89

NOTES: Interrater reliability 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, and 0.81–1.00 is excellent/almost perfect.

Table A.7. Interrater Reliability Kappa or Weighted Kappa for the Staff Assessment of Mental Status Data Elements by Facility Ownership

Data Element	For-Profit (*n* = 369)	Nonprofit (*n* = 135)
Short-term memory OK (b3a)	0.86	0.30
Long-term memory OK (b3b)	0.87	0.69
Is the patient normally able to recall: current season (b3c1)	0.95	0.80
Is the patient normally able to recall: location of own room (b3c2)	0.95	0.87
Is the patient normally able to recall: staff names and faces (b3c3)	0.92	0.80
Is the patient normally able to recall: that he or she is in a care facility (b3c4)	0.94	0.76
Ability to make decisions regarding everyday tasks (b3d)	0.97	0.66

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, and 0.81–1.00 is excellent/almost perfect.

Table A.8. Interrater Reliability Kappa or Weighted Kappa for the Staff Assessment of Mental Status Data Elements by Facility Size

Data Element	Below Setting-Type Median (*n* = 248)	Above Setting-Type Median (*n* = 257)
Short-term memory OK (b3a)	0.81	0.67
Long-term memory OK (b3b)	0.85	0.79
Is the patient normally able to recall: current season (b3c1)	0.97	0.85
Is the patient normally able to recall: location of own room (b3c2)	0.96	0.91
Is the patient normally able to recall: staff names and faces (b3c3)	0.94	0.84
Is the patient normally able to recall: that he or she is in a care facility (b3c4)	0.91	0.88
Ability to make decisions regarding everyday tasks (b3d)	0.94	0.85

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, and 0.81–1.00 is excellent/almost perfect.

Supplementary Tables for Staff Assessment of Patient/Resident Mood (PHQ-9-OV)

Table A.9. Time to Complete the PHQ-9-OV Data Elements by Urbanicity (minutes)

Time to Complete	Urban (*n* = 457)	Nonurban (*n* = 18)	Overall (*n* = 475)
Mean (SD)	3.5 (1.7)	3.6 (1.8)	3.5 (1.7)

Table A.10. Time to Complete the PHQ-9-OV Data Elements by Region (minutes)

Time to Complete	Northeast (*n* = 88)	South (*n* = 149)	Midwest (*n* = 134)	West (*n* = 104)	Overall (*n* = 475)
Mean (SD)	3.4 (1.9)	3.6 (1.7)	3.1 (1.5)	4.0 (1.8)	3.5 (1.7)

Table A.11. Time to Complete the PHQ-9-OV Data Elements by Facility Ownership (minutes)

Time to Complete	For-Profit (*n* = 338)	Nonprofit (*n* = 135)	Overall (*n* = 475)
Mean (SD)	3.3 (1.6)	3.9 (2.0)	3.5 (1.7)

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

Table A.12. Time to Complete the PHQ-9-OV Data Elements by Facility Size (minutes)

Time to Complete	Below Setting-Type Median (*n* = 249)	Above Setting-Type Median (*n* = 225)	Overall (*n* = 475)
Mean (SD)	3.5 (1.7)	3.5 (1.8)	3.5 (1.7)

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

Table A.13. Interrater Reliability Kappa or Weighted Kappa for PHQ-9-OV Data Elements by Urbanicity

Data Element	Urban (*n* = 468)	Nonurban (*n* = 19)
Symptom presence and frequency: little interest or pleasure (e4a)	0.97	0.88
Symptom presence and frequency: feeling down, depressed, hopeless (e4b)	0.98	1.00
Symptom presence and frequency: too little/too much sleep (e4c)	0.98	1.00
Symptom presence and frequency: tired/no energy (e4d)	0.98	1.00
Symptom presence and frequency: poor appetite or overeating (e4e)	0.94	1.00
Symptom presence and frequency: feel bad about self (e4f)	—	—
Symptom presence and frequency: trouble concentrating (e4g)	0.93	1.00
Symptom presence and frequency: moving or speaking slowly (e4h)	0.82	1.00
Symptom presence and frequency: suicidal thoughts (e4i)	—	—
Symptom presence and frequency: short-tempered (e4j)	0.97	1.00
Sum of all symptom frequencies (PHQ-9)[a]	0.94	1.00

NOTES: Interrater reliability 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, and 0.81–1.00 is excellent/almost perfect.
[a] As classified into the three categories shown in Table 4.1 (no depression, minor depression, major depression).

Table A.14. Interrater Reliability Kappa or Weighted Kappa for PHQ-9-OV Data Elements by Region

Data Element	Northeast (*n* = 59)	South (*n* = 160)	Midwest (*n* = 130)	West (*n* = 138)
Symptom presence and frequency: little interest or pleasure (e4a)	—	0.97	0.98	0.96
Symptom presence and frequency: feeling down, depressed, hopeless (e4b)	—	0.99	1.00	0.95
Symptom presence and frequency: too little/too much sleep (e4c)	0.93	0.97	1.00	0.99
Symptom presence and frequency: tired/no energy (e4d)	0.88	0.99	1.00	0.96
Symptom presence and frequency: poor appetite or overeating (e4e)	0.95	0.93	1.00	0.90
Symptom presence and frequency: feel bad about self (e4f)	—	—	—	0.98
Symptom presence and frequency: trouble concentrating (e4g)	0.89	0.91	0.98	0.93
Symptom presence and frequency: moving or speaking slowly (e4h)	0.79	0.88	0.98	0.95
Symptom presence and frequency: suicidal thoughts (e4i)	—	—	—	—
Symptom presence and frequency: short-tempered (e4j)	1.00	0.94	0.99	1.00
Sum of all symptom frequencies (PHQ-9)[a]	0.84	0.89	0.99	0.97

NOTES: Interrater reliability 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, and 0.81–1.00 is excellent/almost perfect.
[a] As classified into the three categories shown in Table 4.1 (no depression, minor depression, major depression).

Table A.15. Interrater Reliability Kappa or Weighted Kappa for PHQ-9-OV Data Elements by Facility Ownership

Data Element	For-Profit (*n* = 354)	Nonprofit (*n* = 131)
Symptom presence and frequency: little interest or pleasure (e4a)	0.97	0.95
Symptom presence and frequency: feeling down, depressed, hopeless (e4b)	0.99	0.96
Symptom presence and frequency: too little/too much sleep (e4c)	0.99	0.95
Symptom presence and frequency: tired/no energy (e4d)	0.98	0.97
Symptom presence and frequency: poor appetite or overeating (e4e)	0.96	0.90
Symptom presence and frequency: feel bad about self (e4f)	—	—
Symptom presence and frequency: trouble concentrating (e4g)	0.96	0.88
Symptom presence and frequency: moving or speaking slowly (e4h)	0.94	0.87
Symptom presence and frequency: suicidal thoughts (e4i)	—	—
Symptom presence and frequency: short-tempered (e4j)	0.98	0.96
Sum of all symptom frequencies (PHQ-9)[a]	0.97	0.96

NOTES: Interrater reliability 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, and 0.81–1.00 is excellent/almost perfect.
[a] As classified into the three categories shown in Table 4.1 (no depression, minor depression, major depression).

Table A.16. Interrater Reliability Kappa or Weighted Kappa for PHQ-9-OV Data Elements by Facility Size

Data Element	Below Setting-Type Median (*n* = 248)	Above Setting-Type Median (*n* = 238)
Symptom presence and frequency: little interest or pleasure (e4a)	0.95	0.99
Symptom presence and frequency: feeling down, depressed, hopeless (e4b)	0.99	0.97
Symptom presence and frequency: too little/too much sleep (e4c)	0.98	0.99
Symptom presence and frequency: tired/no energy (e4d)	0.99	0.96
Symptom presence and frequency: poor appetite or overeating (e4e)	0.93	0.96
Symptom presence and frequency: feel bad about self (e4f)	—	—
Symptom presence and frequency: trouble concentrating (e4g)	0.94	0.93
Symptom presence and frequency: moving or speaking slowly (e4h)	0.93	0.92
Symptom presence and frequency: suicidal thoughts (e4i)	—	—
Symptom presence and frequency: short-tempered (e4j)	0.96	0.98
Sum of all symptom frequencies (PHQ-9)[a]	0.99	0.94

NOTES: Interrater reliability 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, and 0.81–1.00 is excellent/almost perfect.
[a] As classified into the three categories shown in Table 4.1 (no depression, minor depression, major depression).

Supplementary Tables for Observational Assessment of Pain or Distress

Table A.17. Time to Complete the Observational Assessment of Pain or Distress Data Elements by Urbanicity (minutes)

Time to Complete	Urban (*n* = 481)	Nonurban (*n* = 19)	Overall (*n* = 500)
Mean (SD)	2.4 (1.7)	2.4 (1.4)	2.4 (1.7)

Table A.18. Time to Complete the Observational Assessment of Pain or Distress Data Elements by Region (minutes)

Time to Complete	Northeast (*n* = 90)	South (*n* = 161)	Midwest (*n* = 136)	West (*n* = 113)	Overall (*n* = 500)
Mean (SD)	2.6 (1.9)	2.5 (1.6)	2.1 (1.5)	2.5 (1.7)	2.4 (1.7)

Table A.19. Time to Complete the Observational Assessment of Pain or Distress Data Elements by Facility Ownership (minutes)

Time to Complete	For-Profit (*n* = 352)	Nonprofit (*n* = 146)	Overall (*n* = 500)
Mean (SD)	2.3 (1.7)	2.6 (1.7)	2.4 (1.7)

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 Observational Assessment of Pain or Distress Data Elements by Facility Size (minutes)

Time to Complete	Below Setting-Type Median (*n* = 260)	Above Setting-Type Median (*n* = 239)	Overall (*n* = 500)
Mean (SD)	2.5 (1.6)	2.4 (1.8)	2.4 (1.7)

NOTE: Patient/resident numbers in above and below 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 Observational Assessment of Pain or Distress Data Elements by Urbanicity

Data Element	Urban (*n* = 524)	Nonurban (*n* = 19)
Observed pain or distress: nonverbal sounds (d7)	0.79	1.00
Observed pain or distress: vocal complaints (d7)	0.85	1.00
Observed pain or distress: facial expressions (d7)	0.82	0.87
Observed pain or distress: body movements or postures (d7)	0.82	1.00
Observed pain or distress: none observed or documented (d7)	0.88	1.00

Data Element	Urban (*n* = 524)	Nonurban (*n* = 19)
Frequency of patient complaints or evidence of pain in past three days (d8)	0.86	0.88
Did indicators of pain resolve/diminish with pain medications or treatment (d9)	0.90	1.00

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, and 0.81–1.00 is excellent/almost perfect.

Table A.22. Interrater Reliability Kappa or Weighted Kappa for Observational Assessment of Pain or Distress Data Elements by Region

Data Element	Northeast (*n* = 91)	South (*n* = 170)	Midwest (*n* = 136)	West (*n* = 146)
Observed pain or distress: nonverbal sounds (d7)	0.40	0.88	0.88	0.80
Observed pain or distress: vocal complaints (d7)	0.73	0.91	0.96	0.77
Observed pain or distress: facial expressions (d7)	0.72	0.82	0.96	0.75
Observed pain or distress: body movements or postures (d7)	0.64	0.82	0.92	0.82
Observed pain or distress: none observed or documented (d7)	0.77	0.90	0.98	0.85
Frequency of patient complaints or evidence of pain in past three days (d8)	0.63	0.88	0.91	0.87
Did indicators of pain resolve/diminish with pain medications or treatment (d9)	—	0.82	1.00	0.88

NOTES: Interrater reliability 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, and 0.81–1.00 is excellent/almost perfect.

Table A.23. Interrater Reliability Kappa or Weighted Kappa for Observational Assessment of Pain or Distress Data Elements by Facility Ownership

Data Element	For-Profit (*n* = 393)	Nonprofit (*n* = 148)
Observed pain or distress: nonverbal sounds (d7)	0.83	0.72
Observed pain or distress: vocal complaints (d7)	0.84	0.89
Observed pain or distress: facial expressions (d7)	0.84	0.77
Observed pain or distress: body movements or postures (d7)	0.83	0.84
Observed pain or distress: none observed or documented (d7)	0.91	0.84
Frequency of patient complaints or evidence of pain in past three days (d8)	0.91	0.71
Did indicators of pain resolve/diminish with pain medications or treatment (d9)	0.87	1.00

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, and 0.81–1.00 is excellent/almost perfect.

Table A.24. Interrater Reliability Kappa or Weighted Kappa for Observational Assessment of Pain or Distress Data Elements by Facility Size

Data Element	Below Setting-Type Median (*n* = 265)	Above Setting-Type Median (*n* = 277)
Observed pain or distress: nonverbal sounds (d7)	0.85	0.77
Observed pain or distress: vocal complaints (d7)	0.91	0.80
Observed pain or distress: facial expressions (d7)	0.89	0.76
Observed pain or distress: body movements or postures (d7)	0.89	0.77
Observed pain or distress: none observed or documented (d7)	0.92	0.86
Frequency of patient complaints or evidence of pain in past three days (d8)	0.90	0.83
Did indicators of pain resolve/diminish with pain medications or treatment (d9)	0.90	0.91

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, and 0.81–1.00 is excellent/almost perfect.

References

Abdulla, Aza, Nicola Adams, Margaret Bone, Alison M. Elliott, Jean Gaffin, Derek Jones, Roger Knaggs, Denis Martin, Liz Sampson, and Pat Schofield, "Guidance on the Management of Pain in Older People," *Age and Ageing*, Vol. 42, Supp. 1, March 2013, pp. i1–i57.

American Geriatrics Society Panel on the Pharmacological Management of Persistent Pain in Older Persons, "Pharmacological Management of Persistent Pain in Older Persons," *Journal of the American Geriatrics Society*, Vol. 57, No. 8, August 2009, pp. 1331–1346.

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.

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.

Blazer, Dan, *Depression in Late Life*, 3rd ed., New York: Springer Publishing, 2002.

Blytt, Kjersti Marie, Bjørn Bjorvatn, Bettina Husebo, and Elisabeth Flo, "Effects of Pain Treatment on Sleep in Nursing Home Patients with Dementia and Depression: A Multicenter Placebo-Controlled Randomized Clinical Trial," *International Journal of Geriatric Psychiatry*, Vol. 33, No. 4, April 2018, pp. 663–670.

Brenner, Ingrid, and A. Marsella, "Factors Influencing Exercise Participation by Clients in Long-Term Care," *Perspectives*, Vol. 32, No. 4, December 2008, pp. 5–11.

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 Polityńska, 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.

Casten, Robin, M. Powell Lawton, Patricia A. Parmelee, and Morton H. Kleban, "Psychometric Characteristics of the Minimum Data Set I: Confirmatory Factor Analysis," *Journal of the American Geriatrics Society*, June 1998, Vol. 46, No. 6, pp. 726–735.

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.

Charney, D. S., C. F. Reynolds, L. Lewis, B. D. Lebowitz, T. Sunderland, G. S. Alexopoulos, D. G. Blazer, I. R. Katz, B. S. Meyers, P. A. Arean, S. Borson, C. Brown, M. L. Bruce, C. M. Callahan, M. E. Charlson, Y. Conwell, B. N. Cuthbert, D. P. Devanand, M. J. Gibson, G. L. Gottlieb, K. R. Krishnan, S. K. Laden, C. G. Lyketsos, B. H. Mulsant, G. Niederehe, J. T. Olin, D. W. Oslin, J. Pearson, T. Persky, B. G. Pollock, S. Raetzman, M. Reynolds, C. Salzman, R. Schulz, T. L. Schwenk, E. Scolnick, J. Unutzer, M. M. Weissman, R. C. Young, and Depression and Bipolar Support Alliance, "Depression and Bipolar Support Alliance Consensus Statement on the Unmet Needs in Diagnosis and Treatment of Mood Disorders in Late Life," *Archives of General Psychiatry*, Vol. 60, No. 7, July 2003, pp. 664–672.

Chin, Raymond Ping-Hong, Chin-Hung Ho, and Lydia Po-Chee Cheung, "Scheduled Analgesic Regimen Improves Rehabilitation After Hip Fracture Surgery," *Clinical Orthopaedics and Related Research*, Vol. 471, No. 7, July 2013, pp. 2349–2360.

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

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

Cronin-Stubbs, Diane, Carlos F. Mendes de Leon, Laurel A. Beckett, Terry S. Field, Robert J. Glynn, and Denis A. Evans, "Six-Year Effect of Depressive Symptoms on the Course of Physical Disability in Community-Living Older Adults," *Archives of Internal Medicine*, Vol. 160, No. 20, November 2000, pp. 3074–3080.

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.

Dahlhamer, James, Jacqueline Lucas, Carla Zelaya, Richard Nahin, Sean Mackey, Lynn DeBar, Robert Kerns, Michael Von Korff, Linda Porter, and Charles Helmick, "Prevalence of Chronic Pain and High-Impact Chronic Pain Among Adults—United States, 2016," *Morbidity and Mortality Weekly Report*, Vol. 67, No. 36, September 14, 2018, pp. 1001–1006.

Diefenbach, Gretchen J., David F. Tolin, and Christina M. Gilliam, "Impairments in Life Quality Among Clients in Geriatric Home Care: Associations with Depressive and Anxiety Symptoms," *International Journal of Geriatric Psychiatry*, Vol. 27, No. 8, August 2012, pp. 828–835.

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 20, 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, 2018. As of June 20, 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, 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 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

Ell, Kathleen, Jurgen Unützer, Maria Aranda, Kathleen Sanchez, and Pey-Jiuan Lee, "Routine PHQ-9 Depression Screening in Home Health Care: Depression Prevalence, Clinical and Treatment Characteristics and Screening Implementation," *Home Health Care Services Quarterly*, Vol. 24, No. 4, 2006, pp. 1–19.

Eslami, Vahid, Molly E. Zimmerman, Trishdeep Grewal, Mindy Katz, and Richard B. Lipton, "Pain Grade and Sleep Disturbance in Older Adults: Evaluation the Role of Pain, and Stress for Depressed and Non-Depressed Individuals," *International Journal of Geriatric Psychiatry*, Vol. 31, No. 5, May 2016, pp. 450–457.

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.

Garrison, Christopher, Janine Overcash, and Susan McMillan, "Predictors of Quality of Life in Elderly Hospice Patients with Cancer," *Journal of Hospice and Palliative Nursing*, Vol. 13, No. 5, September 2011, pp. 288–297.

Harris, Yael, and James K. Cooper, "Depressive Symptoms in Older People Predict Nursing Home Admission," *Journal of the American Geriatrics Society*, Vol. 54, No. 4, April 2006, pp. 593–597.

Heisel, Marnin J., Paul R. Duberstein, Jeffrey M. Lyness, and Mitchell D. Feldman, "Screening for Suicide Ideation Among Older Primary Care Patients," *Journal of the American Board of Family Medicine*, Vol. 23, No. 2, March 2010, pp. 260–269.

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.

Hyer, Lee, "Depression in Long-Term Care," *Clinical Psychology: Science and Practice*, Vol. 12, No. 3, September 2005, pp. 280–299.

Ishii, Shinya, Joel E. Streim, and Debra Saliba, "Potentially Reversible Resident Factors Associated with Rejection of Care Behaviors," *Journal of the American Geriatrics Society*, Vol. 58, No. 9, September 2010, pp. 1693–1700.

———, "A Conceptual Framework for Rejection of Care Behaviors: Review of Literature and Analysis of Role of Dementia Severity," *Journal of the American Medical Directors Association*, Vol. 13, No. 1, January 2012, pp. 11–23, e1–e2.

Jones, Richard N., Edward R. Marcantonio, and Terry Rabinowitz, "Prevalence and Correlates of Recognized Depression in U.S. Nursing Homes," *Journal of the American Geriatrics Society*, Vol. 51, No. 10, October 2003, pp. 1404–1409.

Kane, Katherine D., Brian P. Yochim, and Peter A. Lichtenberg, "Depressive Symptoms and Cognitive Impairment Predict All-Cause Mortality in Long-Term Care Residents," *Psychology and Aging*, Vol. 25, No. 2, June 2010, pp. 446–452.

Katon, Wayne J., Elizabeth Lin, Joan Russo, and Jürgen Unützer, "Increased Medical Costs of a Population-Based Sample of Depressed Elderly Patients," *Archives of General Psychiatry*, Vol. 60, No. 9, September 2003, pp. 897–903.

Kim, Scott Y. H., Jason H. T. Karlawish, and Eric D. Caine, "Current State of Research on Decision-Making Competence of Cognitively Impaired Elderly Persons," *American Journal of Geriatric Psychiatry*, Vol. 10, No. 2, March–April 2002, pp. 151–165.

Kroenke, Kurt, Dale Theobald, Jingwei Wu, Julie K. Loza, Janet S. Carpenter, and Wanzhu Tu, "The Association of Depression and Pain with Health-Related Quality of Life, Disability, and Health Care Use in Cancer Patients," *Journal of Pain and Symptom Management*, Vol. 40, No. 3, September 2010, pp. 327–341.

Lapane, Kate L., Brian J. Quilliam, Wing Chow, and Myoung Kim, "The Association Between Pain and Measures of Well-Being Among Nursing Home Residents," *Journal of the American Medical Directors Association*, Vol. 13, No. 4, May 2012, pp. 344–349.

Lawton, M. Powell, Robin Casten, Patricia A. Parmelee, Kimberly Van Haitsma, Julie Corn, and Morton H. Kleban, "Psychometric Characteristics of the Minimum Data Set II: Validity," *Journal of the American Geriatrics Society*, Vol. 46, No. 6, June 1998, pp. 736–744.

Leone, Andres F., Francesco Standoli, and Victor Hirth, "Implementing a Pain Management Program in a Long-Term Care Facility Using a Quality Improvement Approach," *Journal of the American Medical Directors Association*, Vol. 10, No. 1, January 2009, pp. 67–73.

Lépine, Jean-Pierre, and Mike Briley, "The Increasing Burden of Depression," *Neuropsychiatric Disease and Treatment*, Vol. 7, Supp. 1, 2011, pp. 3–7.

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.

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.

Murtaugh, Christopher M., Timothy R. Peng, Stanley Moore, and Gill A. Maduro, *Assessing Home Health Care Quality for Post-Acute and Chronically Ill Patients: Final Report*, Washington, D.C.: U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation, August 2008.

Payne, Jennifer L., Jeannie-Marie E. Sheppard, Martin Steinberg, Andrew Warren, Alva Baker, Cynthia Steele, Jason Brandt, and Constantine G. Lyketsos, "Incidence, Prevalence, and Outcomes of Depression in Residents of a Long-Term Care Facility with Dementia," *International Journal of Geriatric Psychiatry*, Vol. 17, No. 3, March 2002, pp. 247–253.

PHQ—*See* Robert L. Spitzer, Kurt Kroenke, and Janet B. Williams, "Validation and Utility of a Self-Report VERSION of PRIME-MD: The PHQ Primary Care Study," *Journal of the American Medical Association*, Vol. 282, No. 18, 1999, pp. 1737–1744.

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 20, 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 20, 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

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.

Rosano, Caterina, Eleanor Marie Simonsick, Tamara B. Harris, Steven B. Kritchevsky, Jennifer Brach, Marjolein Visser, Kristine Yaffe, and Anne B. Newman, "Association Between Physical and Cognitive Function in Healthy Elderly: The Health, Aging and Body Composition Study," *Neuroepidemiology*, Vol. 24, No. 1-2, 2005, pp. 8–14.

Ruo, Bernice, John S. Rumsfeld, Mark A. Hlatky, Haiying Liu, Warren S. Browner, and Mary A. Whooley, "Depressive Symptoms and Health-Related Quality of Life: The Heart and Soul Study," *JAMA*, Vol. 290, No. 2, July 9, 2003, pp. 215–221.

Saliba, Debra, and Joan Buchanan, *Development and Validation of a Revised Nursing Home Assessment Tool: MDS 3.0*, Baltimore, Md.: Centers for Medicare & Medicaid Services, April 2008.

Saliba, Debra, Suzanne DiFilippo, Maria Orlando Edelen, Kurt Kroenke, Joan Buchanan, and Joel Streim, "Testing the PHQ-9 Interview and Observational Versions (PHQ-9 OV) for MDS 3.0," *Journal of the American Medical Directors Association*, Vol. 13, No. 7, September 2012, pp. 618–625.

Schoevers, Robert A., Mirjam I. Geerlings, Aartjan T. F. Beekman, Brenda W. Penninx, Dorly J. H. Deeg, Cees Jonker, and Willem van Tilburg, "Association of Depression and Gender with Mortality in Old Age: Results from the Amsterdam Study of the Elderly (AMSTEL)," *British Journal of Psychiatry*, Vol. 177, No. 4, October 2000, pp. 336–342.

Shen, Xian, Ilene H. Zuckerman, Jacqueline B. Palmer, and Bruce Stuart, "Trends in Prevalence for Moderate-to-Severe Pain and Persistent Pain Among Medicare Beneficiaries in Nursing Homes, 2006–2009," *Journals of Gerontology: Series A*, Vol. 70, No. 5, May 2015, pp. 598–603.

Slaughter, Susan E., Misha Eliasziw, Debra Morgan, and Neil Drummond, "Incidence and Predictors of Excess Disability in Walking Among Nursing Home Residents with Middle-Stage Dementia: A Prospective Cohort Study," *International Psychogeriatrics*, Vol. 23, No. 1, February 2011, pp. 54–64.

Spitzer, Robert L., Kurt Kroenke, and Janet B. Williams, "Validation and Utility of a Self-Report VERSION of PRIME-MD: The PHQ Primary Care Study," *Journal of the American Medical Association*, Vol. 282, No. 18, 1999, pp. 1737–1744.

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.

Sullivan-Singh, Sarah J., Kathryn Sawyer, Dawn M. Ehde, Kathleen R. Bell, Nancy Temkin, Sureyya Dikmen, Rhonda M. Williams, and Jeanne M. Hoffman, "Comorbidity of Pain and Depression Among Persons with Traumatic Brain Injury," *Archives of Physical Medicine and Rehabilitation*, Vol. 95, No. 6, June 2014, pp. 1100–1105.

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.

Teresi, Jeanne, Robert C. Abrams, Douglas Holmes, Mildred M. Ramirez, Joseph Eimicke, "Prevalence of Depression and Depression Recognition in Nursing Homes," *Social Psychiatry and Psychiatric Epidemiology*, Vol. 36, No. 12, December 2001, pp. 613–620.

U.S. Department of Agriculture, "Documentation: 2010 Rural-Urban Commuting Area (RUCA) Codes," webpage, October 12, 2016. As of June 20, 2019: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/documentation/

Wang, Li, Gerald Van Belle, Walter B. Kukull, and Eric B. Larson, "Predictors of Functional Change: A Longitudinal Study of Nondemented People Aged 65 and Older," *Journal of the American Geriatric Society*, Vol. 50, No. 9, September 2002, pp. 1525–1534.

Zanca, Jeanne M., Marcel P. Dijkers, Flora M. Hammond, and Susan D. Horn, "Pain and Its Impact on Inpatient Rehabilitation for Acute Traumatic Spinal Cord Injury: Analysis of Observational Data Collected in the SCIRehab Study," *Archives of Physical Medicine and Rehabilitation*, Vol. 94, No. 4, Supplement, April 2013, pp. S137–S144.

Ziegelstein, Roy C., "Depression in Patients Recovering from a Myocardial Infarction," *JAMA*, Vol. 286, No. 13, October 3, 2001, pp. 1621–1627.