Development and Evaluation of Candidate Standardized Patient Assessment Data Elements: Findings from the National Beta Test (Volume 5: 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-3004/4-CMS

Development and Evaluation of Candidate Standardized Patient Assessment Data Elements:

Findings from the National Beta Test (Volume 5: 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/RR3004Z4

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

Preface

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

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

Contents

Figures and Tables

Figures

Tables

Abbreviations

ADLs	activities of daily living
BIMS	Brief Interview for Mental Status
CARE	Continuity Assessment Record and Evaluation
CMS	Centers for Medicare & Medicaid Services
CY	calendar year
DIF	differential item functioning
ECV	explained common variance
FY	fiscal year
HHA	home health agency
IMPACT	Improving Medicare Post-Acute Care Transformation
IRF	inpatient rehabilitation facility
IRF-PAI	Inpatient Rehabilitation Facility Patient Assessment Instrument
IRT	item response theory
LCDS	Long-Term Care Hospital CARE Data Set
LTCH	long-term care hospital
MDS	Minimum Data Set
NIH	National Institutes of Health
OASIS	Outcome and Assessment Information Set
PAC	post-acute care
PAC-PRD	Post-Acute Care Payment Reform Demonstration
PHQ	Patient Health Questionnaire
PROMIS	Patient Reported Outcome Measurement Information System
SD	standard deviation
SME	subject-matter expert
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 5 of the final report on the National Beta Test, which includes the identification and testing of candidate SPADEs in the clinical categories of *mental status* 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, which included two Alpha feasibility tests held in select CMS regions,[1] two technical expert panels (TEPs),[2] two subregulatory calls for public comment,[3] and one notice of proposed rulemaking for the Fiscal Year (FY)/Calendar Year (CY) 2018 proposed rules.[4] The results of these activities informed the content and design of the National Beta Test.

The National Beta Test included data collection within 143 PAC facilities/agencies across 14 markets in the United States (listed in Volume 2 of the final report[5]), from November 2017 to August 2018. The overarching goal of the National Beta Test was to evaluate the feasibility, reliability, and validity of candidate SPADEs to identify a subset of data elements for standardization across PAC settings. Candidate SPADEs were considered if they met the requirements of being feasible, clinically useful, and having the potential to improve quality. Trained research nurses and/or staff at participating PAC facilities/agencies administered all National Beta Test assessment protocols. A subset of National Beta Test assessments was 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, 2017b, 2017c, 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 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 of each data element for all admission data, first combined across settings (overall) and then by setting.
- Statistical comparison of versions for the Patient-Reported Outcomes Measurement Information System (PROMIS) Depression, PROMIS Anxiety, and Pain Interview data elements.[7]
- Evaluation of differential item functioning (DIF) according to version, setting, gender, and age for PROMIS Depression and PROMIS Anxiety. DIF occurs when items do not

[6] Edelen et al., 2019b.

[7] Dewitt et al., 2018. *PROMIS*, *Patient-Reported Outcomes Measurement Information System*, and the PROMIS logo are marks owned by the U.S. Department of Health and Human Services.

assess the underlying construct equally well, or vary in meaning, between different subgroups of patients/residents.
- Extent of missing data for each data element overall. Missing data were minimal and did not vary by setting and so are only briefly summarized.
- Average time to complete the assessment of each data element, for each data element overall and by setting.

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 follow 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.
- Item and scale reliability for the PROMIS Depression and Anxiety scales included average inter-item correlations, item-total correlations, and Cronbach's alpha. Item-total correlations of $r < 0.4$ suggest that items should be considered for deletion in a scale.

Validity

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

- For PROMIS Depression and Anxiety, explained common variance (ECV)[8] was calculated based on an exploratory bifactor model,[9] which estimates one general factor for all the items and allows for the presence of additional specific factors to capture unexplained residual associations among items. The ECV is the ratio of the common variance explained by the general factor as compared with the specific factors and has a maximum value of 1.0. ECV values greater than or equal to 0.60 may indicate that items in a scale are sufficiently unidimensional and are thus assessing a single construct. Although the dimensionality of PROMIS items has been established, the items have not been used in PAC settings, so these analyses were conducted to confirm unidimensionality for the population of PAC patients and residents.

Stability/Change over Time

- Comparison of admission and discharge frequency data for each data element overall and by setting.
- Degree of change in rates or scores depending on the day a patient/resident was assessed after admission (i.e., Day 1), within the Day 3, 5, and 7 repeat assessment design. These results are reported for all data elements included in the repeat assessment design overall and by setting.

Sensitivity to National Representativeness

- Sensitivity analyses for each data element to confirm that performance does not vary according to urbanicity as classified by rural-urban commuting area codes (metropolitan and micropolitan [urban] versus small town and rural [nonurban]),[10] 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 data elements evaluated in this volume. 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.

Statistical Tests

- Categorical associations were statistically evaluated using chi-square tests of independence and, in the case of ordinal data, Mantel-Haenszel chi-square.[11] Significant results from chi-square tests are reported in the following format: ($\chi^2_{(df)} = X.X$, $p < 0.05$), where *df* are degrees of freedom and *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).

---

[8] Rodriguez, Reise, and Haviland, 2016a; Rodriguez, Reise, and Haviland, 2016b.

[9] Reise, 2012.

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

[11] Mantel and Haenszel, 1959.

- 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: $F_{(df)} = X.X$, $p < 0.001$ or $t_{(df)} = X.X$, $p < 0.001$, where *df* are degrees of freedom and *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 the amount of time it takes to complete assessments).
- Associations involving two continuous variables were statistically evaluated using a Pearson correlation (*r*) to determine whether a linear association emerged between one continuous variable (e.g., length of stay) and another continuous variable (e.g., PHQ-9 total score). Significant results are reported in the following format: Pearson correlation $r = 0.XX$, $p < 0.05$, where *X*'s are the numerical values corresponding to the magnitude of the linear association. Values can be positive (i.e., scores on both variables increase or decrease together) or negative (i.e., scores on one variable increase while scores on the other variable decrease).
- We used the chi-squared likelihood-ratio statistic as the DIF detection criteria (alpha < 0.01) and the pseudo McFadden R^2 of magnitude (≤ 0.02) in model comparisons. The DIF detection approach is sensitive and often identifies DIF that is very small in magnitude. This 0.02 value for R^2 magnitude is conventionally used in the development of PROMIS instruments to identify statistically significant DIF that is large enough to warrant further examination.
- Effect sizes for many of the significant findings are reported using Cohen's *d* to further characterize the importance of statistically significant findings.[12] 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,[13] where each significance test is evaluated against an adjusted critical value. We set our desired level of significance at 0.01 to minimize Type I error and increase confidence in significant effects.

[12] Cohen, 2013.

[13] Benjamini and Hochberg, 1995.

2. Standardized Assessment of Mental Status and Pain

As described in Chapter 1, candidate SPADEs were identified for inclusion in the National Beta Test through a series of activities that included information gathering, stakeholder outreach, and Alpha field testing. This chapter provides background on the importance of standardized assessment in the clinical categories of *mental status* and *pain*, the activities undertaken to identify candidate SPADEs for these categories during the project period, and the final data elements tested in the National Beta Test. This chapter also gives an overview of the results presented in subsequent chapters of this volume.

Mental Status: Depression

Depression is a common mental health problem in older adults and is particularly common in PAC settings. Estimated rates of clinical depression range from 9 to 28 percent in HHAs[14] and 6 to 45 percent in SNFs,[15] but depression generally is thought to be under-evaluated and under-detected in PAC settings.[16] The deleterious effects of depression on quality of life,[17] physical function, [18] pain,[19] rejection of care behaviors,[20] and increased mortality from other causes have been well documented.[21] Depression is also associated with worse outcomes with chronic illness;[22] slower recovery rates following acute medical events, such as hip fracture,[23] stroke,[24] myocardial infarction,[25] and delirium;[26] and increased risk of institutionalization.[27] Depression

[14] Ell et al., 2006; Bruce et al., 2002.

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

[16] Teresi et al., 2001; Brown et al., 2004; Brown, Lapane, and Luisi, 2002; Shao et al., 2011.

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

[18] Slaughter et al., 2011.

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

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

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

[22] Rapp et al., 2011; Ziegelstein, 2001; Carson et al., 2003; DiMatteo, Lepper, and Croghan, 2000; Katon et al., 2004; Morley, 2007.

[23] Givens, Sanft, and Marcantonio, 2008; Kamholz and Unützer, 2007.

[24] Ayerbe et al., 2013; Robinson, 2003.

[25] Ziegelstein, 2001.

[26] Givens, Jones, and Inouye, 2009.

[27] Luppa et al., 2010; Samus et al., 2009; Sheeran, Byers, and Bruce, 2010; Young, 2009.

screeners help PAC providers better understand the needs of their patients and residents by prompting further evaluation, 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 their stay, and identifying appropriate ongoing treatment and support needs at the time of discharge.

Information Gathering

We surveyed the existing data elements through a web-based search and literature review to identify those related to depression assessment. We also examined current PAC instruments and reviewed the data elements tested as part of the Post-Acute Care Payment Reform Demonstration (PAC-PRD).

We identified two depression screeners used to assess the signs and symptoms of depression in the existing PAC instruments: the nine-item Patient Health Questionnaire (PHQ-9),[28] which assesses each of the criteria for major depressive disorder outlined in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5),[29] and a two-item version of the PHQ-9 (the PHQ-2), which assesses only the cardinal criteria for depression: depressed mood and anhedonia (i.e., an inability to experience pleasure from activities that are usually enjoyable). The PHQ-9 is currently used in the MDS, and the PHQ-2 is used in OASIS.

The PHQ-9 is freely available in more than 80 translations and has been validated for use with older adults,[30] in home health,[31] in SNFs,[32] and with rehabilitation populations.[33] The PHQ-9 has also demonstrated the ability to identify clinically important signs and symptoms of depression in adults of all ages,[34] and it has been shown to be a reliable and valid screening tool for detecting signs and symptoms of major depressive disorder in patients with complex medical issues, including stroke and traumatic brain injury.[35]

The PHQ-9 was administered to 3,258 residents of SNFs in the national validation of the MDS 3.0, where it was found to be highly correlated with a physician diagnosis of depression in

---

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

[29] American Psychiatric Association, 2013.

[30] Kroenke, Spitzer, and Williams, 2001; Spitzer, Kroenke, and Williams, 1999; Klapow et al., 2002; Löwe et al., 2004.

[31] Blazer, 2002.

[32] Saliba et al., 2012.

[33] Williams et al., 2005.

[34] Rahman and Applebaum, 2009.

[35] Williams et al., 2005; Fann et al., 2005.

both self-report and interview assessment.[36] Four systematic meta-analyses have evaluated the ability of the PHQ-9 assessment of the signs and symptoms of depression to accurately identify patients/residents whose condition received a physician diagnosis of depression in primary care and other outpatient settings, finding it to have good sensitivity with little variation across care settings.[37]

The PHQ-2 is used in OASIS. Efforts to validate the shorter PHQ have been less extensive and have generally not been focused on PAC populations. Among studies conducted in primary care centers with large samples of adults, however, the PHQ-2 has performed adequately both as a screening tool for identifying the symptoms of depression and for assessing depression severity.[38] It has also been shown to be sensitive to change, accurately reflecting improved, unchanged, and deteriorated depression outcomes over time.[39] The PHQ-2 was tested in the PAC-PRD and found to be reliable across the four PAC settings (kappas ranged from 0.74 to 0.91).[40]

In collaboration with the PROMIS team from Northwestern University, we also identified the subset of depression items from the PROMIS Item Bank as candidate items for consideration.[41] The PROMIS Item Bank was developed as part of a National Institutes of Health (NIH) Roadmap initiative that set the standard for modern behavioral health measurement development. PROMIS is at the forefront of NIH efforts to fund research that advances behavioral health measurement by developing new self-reporting tools based on the principles of item response theory (IRT).[42]

We also identified the Short-Form Geriatric Depression Scale (GDS-15)[43] and the Center for Epidemiological Studies Depression Scale (CES-D)[44] as candidate data elements for standardization in PAC settings. Each of these tools is known to have strong psychometric properties.[45] The Hamilton Depression Rating Scale (HAM-D)[46] and the Montgomery-Åsberg

---

[36] Saliba et al., 2012; Fann et al., 2005.

[37] Gilbody et al., 2007; Manea, Gilbody, and McMillan, 2012.; Wittkampf et al., 2007; Moriarty et al., 2015.

[38] Li et al., 2007; Löwe, Kroenke, and Gräfe, 2005.

[39] Löwe, Kroenke, and Gräfe, 2005.

[40] Gage et al., 2012.

[41] Pilkonis et al., 2011.

[42] Dewitt et al., 2018. *PROMIS*, *Patient-Reported Outcomes Measurement Information System*, and the PROMIS logo are marks owned by the U.S. Department of Health and Human Services.

[43] Yesavage and Sheikh, 1986.

[44] Radloff, 1977.

[45] Feher, Larrabee, and Crook, 1992; Roth et al., 2008; Schalet et al., 2016.

[46] Rohan et al., 2016; Reynolds and Kobak, 1995.

Depression Rating Scale (MADRS)[47] were also considered and have demonstrated good psychometric properties but were judged by our clinical advisers to be impractical for PAC use.

Stakeholder Feedback and Field Testing

During the focus groups with PAC staff, several providers raised questions about how important or advisable it is to strictly adhere to the wording of the depression screeners and suggested that alternative words to *depression* might mitigate patient/resident fears of being negatively labeled. They also emphasized the importance of distinguishing situational depression arising from being in an institutional setting from major depressive disorder, a comment which pertained less to the assessment process and more to the clinical care required for patients/residents who screen positive. The TEP members also voiced concerns about the time required to administer mood assessments and resulting assessor burden.

PROMIS Depression (which contains a subset of depression-related items from the PROMIS Item Bank), the PHQ-9, the PHQ-2, the GDS-15, and the CES-D were presented at the first convening of the TEP for consideration and rating.[48] Members of the TEP affirmed the importance of screening for depressed mood in PAC settings, noting that patients/residents who are experiencing depression may be less able or willing to fully participate in their care. The TEP members also discussed the general challenges of depression screening in PAC, including possible assessor discomfort with patient interview questions related to depression and the need for facilities to have and communicate a clear plan to staff about what to do when a patient/resident screens positive for possible mood disorder.

During the rating exercise, the PHQ-2 received the highest marks, being rated "very good" across all four dimensions: (1) potential for improving quality, (2) validity and reliability, (3) feasibility for use in PAC, and (4) utility for describing case mix. The PHQ-9 was also rated as "very good" overall and was perceived to have the highest validity and reliability, although it ranked slightly lower on average than the PHQ-2. The remaining items were generally rated as "good," though their feasibility for use in PAC consistently received lower ratings

Although the TEP was satisfied with the reliability, validity, and utility of the PHQ-2, some panelists suggested boosting the sensitivity of the two-item data element with additional items from the PHQ-9. During a subsequent meeting with federal subject-matter experts (SMEs), the group endorsed the proposed "PHQ-2 to 9" gateway approach, which transitions patients/residents that demonstrate signs of depressed mood and anhedonia under the PHQ-2 to the lengthier PHQ-9. Structuring the data element in this fashion ensures that patients/residents with indications of depressive symptoms based on the PHQ-2 receive the longer assessment, while reducing the length of the assessment for those who do not report the cardinal symptoms of depression.

---

[47] Svanborg and Åsberg, 2001.

[48] RAND Corporation, 2017a.

Both the PHQ-2 and PHQ-9 were posted for public comment in 2016, along with a request for feedback on the potential use of the PHQ-2 as a gateway to the PHQ-9. Commenters expressed support for the PHQ-2 to 9 gateway approach, noting that it would balance reporting burden with the ability to collect more in-depth information about symptoms of depression.

The first Alpha feasibility test (Alpha 1) piloted the PHQ-2 to 9, and the gateway data element performed well. It was found highly feasible to administer and minimally time-consuming. The first two items took between one and three minutes to administer, and an additional three minutes were required when it was necessary to administer all nine items. Assessors suggested that additional training on scoring might be needed, but the items had excellent interrater reliability across all settings.

Candidate SPADEs in the National Beta Test

After thorough consideration through the activities described above, the PHQ-2-to-9 and eight selected data elements from the PROMIS Depression Item Bank were included in the National Beta Test. The data elements are presented and results are described in Chapters 3 and 4 of this volume.

Mental Status: Anxiety

Anxiety disorders are common in the general population[49] and in older adults,[50] who represent a high proportion of patients/residents receiving PAC services. In older adults, anxiety is associated with impairment in function[51] and increased risk for other mental health conditions. This is because anxiety can be associated with other psychological and physical distress (e.g., depression, loss of independence, and increased sensitivity to pain),[52] decreased functional status (e.g., resistance to care, decreased desire to participate in ADLs),[53] and poorer outcomes and increased mortality from other conditions.[54] Identifying symptoms of anxiety can lead to follow-up efforts to identify causes and contributing factors and to identify treatment, support, or environmental modifications that could help alleviate symptoms and ensure patient/resident safety.

---

[49] McLean, et al., 2011; Kessler et al., 2005.

[50] Smalbrugge et al., 2005; Préville et al., 2004.

[51] Stein et al., 2005.

[52] Smalbrugge et al., 2005; Diefenbach and Goethe, 2006; Ocañez, McHugh, and Otto, 2010.

[53] Schultz, Hoth, and Buckwalter, 2004.

[54] Carrière et al., 2013.

Information Gathering

In light of the high incidence of anxiety-related distress in PAC patients/residents,[55] we identified a subset of self-report anxiety items from the PROMIS Item Bank as candidates for standardization. The PROMIS Item Bank for anxiety includes 29 items that focus on fear (e.g., fearfulness, feelings of panic), anxious misery (e.g., worry, dread), hyperarousal (e.g., tension, nervousness, restlessness), and somatic symptoms related to arousal (e.g., cardiovascular symptoms, dizziness). Because of the length of the complete PROMIS Item Bank and concerns that all items might not be relevant to the PAC population, we initially selected a reduced list of 19 items for consideration. These items were selected in late 2016 based on feedback from project team advisers, TEP members, and consumer and provider stakeholders regarding their applicability for patients/residents in the four PAC settings.

Stakeholder Feedback and Field Testing

Members of the TEP at the second convening (January 2017) provided general support for the importance of assessing anxiety across PAC settings and agreed that the PROMIS Anxiety Item Bank represented a good resource for this purpose. Results from the TEP members' rating sheets demonstrated good support of the PROMIS Anxiety items across all dimensions (with an average overall score of 3.5 out of 4). As with the other data elements in the mental status assessment category, TEP members were interested in shortening the anxiety item list. The group then discussed the fact that the advantage of the item bank is that the list of items can be shaped to suit the focus needed without compromising the reliability and validity of the data element. Some questioned the value of including items to which a majority of patients/residents would reply "yes," such as those that ask about worry. However, another panelist pointed out that these items do not prompt a yes/no response but rather inquire about frequency. While the TEP panelists generally concurred with the importance of assessing anxiety in PAC patients/residents, some commented that they would like to see evidence of predictive value in these items (e.g., that the collection of symptoms and behaviors in the item bank corresponds to a physician diagnosis of anxiety) before asking patients/residents about it.

PROMIS Anxiety was subsequently posted for public comment, and it received strong support from commenters, who believed it could improve person-centered care and care planning. Commenters expressed that PROMIS Anxiety might be helpful to describe case mix and may be valid, reliable, and feasible to implement. Other support touched on the relevance and meaningfulness (e.g., because anxiety is asked about in a functional context) of PROMIS Anxiety items to patients in all PAC settings and the importance of addressing anxiety that can result in a hospital readmission or emergency room visit.

---

[55] Ayers et al., 2007; Osborn, Mathias, and Fairweather-Schmidt, 2016; Campbell Burton, 2013; Seitz, Purandare, and Conn, 2010.

Some commenters had concerns about PROMIS Anxiety's noninstitutional focus and ambiguity. Other commenters said that using PROMIS Anxiety to establish case mix may not add value or utility. Several commenters questioned PROMIS Anxiety's feasibility, validity, and reliability, and they particularly called for validation in PAC settings. Other concerns included redundancy, burden, ability to evaluate residents unable to self-report anxiety, and the specified look-back periods.

Eleven anxiety items from the PROMIS Item Bank were tested in Alpha 2 feasibility testing. Interrater relatability ranged from 0.8 to 1, with most ratings falling between 0.95 and 1, indicating almost perfect agreement between assessors. The anxiety items were quick to administer, taking three to six minutes to complete. Assessors' comments suggested that the Anxiety items were straightforward to administer and highly consistent across assessors. Additionally, assessors were concerned that the 11 items may be burdensome to the patients/residents.

Candidate SPADEs in the National Beta Test

After thorough consideration through the activities described above and in consultation with the PROMIS development team, TEP members, and clinical advisers, we selected a subset of eight PROMIS Anxiety items for inclusion in the National Beta Test. The data elements are presented and results are described in Chapter 5 of this volume.

Pain

Pain is a common condition among adults of all ages. A Centers for Disease Control and Prevention analysis of 2016 National Health Interview Study data found that 8 percent of Americans report high-impact chronic pain —that is, pain that limits life or work activities on most days or every day in the past six months.[56] Regular pain occurs in 25 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.[57] Pain is frequent among those receiving home health care, with 53 percent reporting daily pain interfering with activity on admission.[58] 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, neuralgia from diabetes mellitus, peripheral vascular disease, and such infections as herpes zoster/shingles.[59]

---

[56] Dahlhamer et al., 2018.

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

[58] Murtaugh et al., 2008.

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

Conditions causing pain in older adults may be associated with depression,[60] sleep disturbance,[61] and lower participation in rehabilitation activities.[62]

Information Gathering

Data elements applicable to the assessment of pain were surveyed for both interview-based data elements and staff observational assessments, which are discussed separately in Volume 8.[63] We first considered the existing PAC instruments and the data elements tested in the PAC-PRD and then conducted a web-based search and literature review to identify possible additional data elements related to pain assessment.

Current PAC assessment instruments vary with respect to pain assessment. The MDS includes both interview- and observation-based pain data elements, with the interview data elements focusing on self-report of the presence of pain, its severity, its effect on sleep, and its effect on activities. The OASIS, IRF-PAI, and LCDS do not include interview-based data elements to assess pain. The PAC-PRD tested interview-based pain assessment data elements that align closely with those in the MDS. These data elements all demonstrated good reliability, indicating they could be implemented across all four PAC settings without further testing.

In addition to the data elements already in use, the broader literature search yielded 42 interview-based data elements for the assessment of pain. Because pain is a subjective experience for which there are no objective biological markers, self-report is often considered to be the gold standard for identifying the presence, location, intensity, and duration of pain.[64] Among the 42 interview-based pain assessment tools found in the literature review, the assessments captured six constructs:

- pain severity (e.g., no pain, worst pain)
- pain interference (e.g., pain made it hard to sleep at night)
- pain frequency (e.g., rarely, almost constantly)
- pain relief (e.g., no relief, complete relief)
- pain description (e.g., throbbing, stabbing)
- pain location (e.g., indicate on drawing of body).

Studies on the effectiveness of these assessments typically compared patient/resident preferences among the Verbal Rating Scale, Visual Analog Scale, and the Numeric Rating Scale.[65] Others collected psychometric data on the PROMIS Short Form, the Geriatric Pain

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

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

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

[63] Edelen et al., 2019c.

[64] Alexander et al., 2005; Teno et al., 2004; Lukas et al., 2013.

[65] Hjermstad et al., 2011; Bahreini, Jalili, and Moradi-Lakeh, 2015.

Measure, the Brief Pain Inventory, the McGill Pain Questionnaire, and the Functional Pain Scale, but each of these had its drawbacks in terms of time needed to complete or lack of testing among adults among the PAC population.[66]

Challenges with pain assessment include the subjectivity of many pain measures and determining an optimal approach for individuals with severe cognitive impairment or inability to communicate wants and needs.[67]

Stakeholder Feedback and Field Testing

As outlined in Volume 2, stakeholders were given multiple opportunities to comment on data elements in various stages of development leading up to the National Beta Test. Participants of focus groups held in June and July of 2018 noted the need to assess pain in terms of its frequency, to assess whether it is acute or chronic, and to consider patient/resident preferences regarding treatment. They also emphasized the need to understand whether pain was being treated or managed. Participants also noted that for pain data elements, the time at which the assessment is conducted—and by whom—could also affect how assessments are coded.

At the first convening of the TEP (in April 2016), members affirmed the feasibility and clinical utility of pain as a concept in a standardized assessment. The TEP agreed that data elements on pain frequency, patients'/residents' perceptions of adequacy of pain relief/pain treatment regimen, and pain interference with ability to participate in therapies versus other activities should be addressed. The TEP members were particularly supportive of the data elements that focused on how pain interferes with activities, because understanding the extent to which pain interferes with function would enable clinicians to determine the need for appropriate pain treatment. There was also general agreement that adding a graded response scale for several items (instead of a simple yes/no) would be feasible and clinically useful for pain assessment. TEP members rated the PAC-PRD version of the data elements highly, particularly with the addition of items on pain frequency, relief, management, and interference with quality of life.

Federal SMEs emphasized the need to determine an appropriate look-back period, with some noting that five days may be too long of a window. Other SMEs noted the need to assess pain considering the patient's/resident's pain management goals, especially for use during care transitions.

Two data elements that assess pain presence and pain severity were initially put forth for public comment. Several commenters said that it is important to capture pain presence and severity in PAC facilities and across the care continuum. Commenters also noted that a standardized set of data elements, assessing multiple dimensions of pain in addition to presence and severity, could help PAC providers assess patient/resident pain uniformly across the continuum of care and improve care for patients.

---

[66] Gloth et al., 2002; Keller et al., 2004; Tan et al., 2004; Ferrell, Stein, and Beck, 2000.

[67] Lichtner et al., 2014; Herr, Bjoro, and Decker, 2006; Zwakhalen et al., 2006.

Pain Interview data elements that address pain presence, frequency, severity, relief, effect on sleep, and interference with therapy and other activities were subsequently included in Alpha 1 feasibility testing. Assessors participating in testing were able to complete the assessment with a very high degree of reliability, and interrater agreement was nearly perfect for all data elements. Assessors generally found the administration of the data elements straightforward, though some noted that patients/residents may need clarification on whether questions on pain presence should be answered with or without medication.

Data elements presented to the TEP at the January 2017 meeting (the second TEP convening) included those in the Alpha 1 phase of feasibility testing, with modification based on prior feedback. The changes were well received, and one TEP member commented that setting the look-back period to three days would be helpful in making the data elements more applicable across settings. The timing of pain assessment was the subject of some debate—twice daily versus "morning" and "evening," which might be interpreted differently in different settings—because opportunities to complete assessments through home health or facilities could vary considerably depending on when daily care is provided. However, the TEP conceded that providing guidance on the precise timing of data collection for assessments is out of scope for CMS.

We also solicited public feedback during the second public comment period (April to June 2017) on Pain Interview data elements that address frequency, severity, relief, effect on sleep, and interference with therapy and other activities. These data elements had been modified based on previous stakeholder input and testing. A few commenters shared their support for assessing pain, the potential for pain assessment to improve the quality of care, and for the validity and reliability of the data elements. Commenters thought the items on pain frequency, severity, and effect on sleep would be suitable for PAC settings, but the frequency and severity items might be redundant with the MDS. Criticisms of the items on pain interference, severity, and relief centered on cross-setting feasibility and utility, particularly the applicability of these items to all PAC settings. For example, one commenter thought that interference items were clearly developed for non-IRF settings and were too simplistic for cross-setting utility. Some commenters also raised general concerns regarding validity, burden, wording, and look-back periods, and some were concerned about drug-seeking behavior and suggested use of a gateway question to address this.

Candidate SPADEs in the National Beta Test

Pain Interview data elements were included in the National Beta Test in two versions with different look-back periods. The data elements are presented and results are described in Chapter 6.

Summary of Candidate SPADEs in National Beta Test

The Mental Status and Pain data elements evaluated in the National Beta Test communicative sample are shown in Table 2.1. This table also lists the evaluative 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. Mental Status and Pain Data Elements Evaluated in the National Beta Test

Data Element	Input Opportunities	National Beta Test Inclusion Notes	Current Assessment Instrument Use
PHQ-2 to 9	Public Comment 1, FY/CY 2018 proposed rule, Alpha 1		PHQ-2 (OASIS) PHQ-9 (MDS)
PROMIS[b] Depression	TEP/stakeholder review	Two versions tested in the National Beta Test[c]	
PROMIS Anxiety	Alpha 2, Public Comment 2	Two versions tested in the National Beta Test[c]	
Pain	Public Comment 1, Alpha 1, Public Comment 2	Two versions tested;[c] included in Day 3, 5, 7 test	Presence (OASIS[a], MDS) Frequency, severity, effect on sleep (MDS) Activities (OASIS, MDS)

[a] Item M1240 Formal Pain Assessment removed from the OASIS-D, effective January 1, 2019.
[b] All PROMIS materials are © 2008–2019 PROMIS Health Organization. PROMIS is a registered trademark of the U.S. Department of Health and Human Services. All modifications made with permission. Information on PROMIS instruments is available at http://www.healthmeasures.net.
[c] Two versions of the PROMIS Depression, PROMIS Anxiety, and Pain data elements were evaluated in different samples. PROMIS Depression and Anxiety data elements were tested using the standard format and with the “past 3 days.” Pain Interview data elements were tested using the “past 3 days” and “past 5 days.”

3. PHQ-2 to 9

Data Element Description

The PHQ-2 and PHQ-9 protocols are established screening tools used by physicians, hospitals, SNFs, HHAs, and other settings to identify signs and symptoms of depression. The PHQ-2 contains the first two of the nine questions included in the PHQ-9. Screening for signs and symptoms of depression is important because undetected depression can lead to degraded physical and mental health and functioning, increased medical care utilization and costs, reduced quality of life, and premature death. The PHQ-2 to 9 uses the two questions in the PHQ-2 as a gateway for the longer PHQ-9 (the PHQ-2 to 9), such that if the patient/resident scores beyond a threshold level indicating signs and symptoms of possible depression from the first two data elements, the assessor continues to administer the remaining seven data elements from the PHQ-9. Patients/residents who report few or no depressive symptoms in the PHQ-2 (i.e., experiencing the first two symptoms either not at all or less than half the days) would not be asked the additional data elements on the PHQ-9. The PHQ-2 to 9 data elements are completed through patient/resident interview. The PHQ-2 is currently used in OASIS-D, and the PHQ-9 is currently used in the MDS. Pfizer Inc. holds the copyright for all the PHQ screeners, including the PHQ-9; Pfizer Inc. has granted permission to use this instrument in association with the PAC assessment instruments. The PHQ-2 to 9 data elements as assessed in the National Beta Test are shown in Figure 3.1.

Figure 3.1. PHQ-2 to 9 Data Elements

E1a1. SYMPTOM PRESENCE
ASK PATIENT/RESIDENT: "Over the last 2 weeks, have you been bothered by little interest or pleasure in doing things?"

- ☐ 0 = No **[SKIP TO E1b1]**
- ☐ 1 = Yes
- ☐ 7 = **Patient/resident declined to respond [SKIP TO E1b1]**
- ☐ 9 = **Unknown or unable to assess [SKIP TO E1b1]**

E1a2. SYMPTOM FREQUENCY
ASK PATIENT/RESIDENT: "Over the last 2 weeks, how often have you been bothered by having 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)
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

E1b1. SYMPTOM PRESENCE
ASK PATIENT/RESIDENT: "Over the last 2 weeks, have you been bothered by feeling down, depressed, or hopeless?"

- ☐ 0 = No **[SKIP TO E1c1]**
- ☐ 1 = Yes
- ☐ 7 = **Patient/resident declined to respond [SKIP TO E1c1]**
- ☐ 9 = **Unknown or unable to assess [SKIP TO E1c1]**

E1b2. SYMPTOM FREQUENCY
ASK PATIENT/RESIDENT: "Over the last 2 weeks, how often have you been bothered by feeling 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)
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

If either E1a2 or E1b2 is coded 2 or 3, CONTINUE asking the questions below. If not, END the PHQ interview and SKIP to E-TIME.

E1c1. SYMPTOM PRESENCE
ASK PATIENT/RESIDENT: "Over the last 2 weeks, have you been bothered by trouble falling or staying asleep, or sleeping too much?"

- ☐ 0 = No **[SKIP to E1d1]**
- ☐ 1 = Yes
- ☐ 7 = **Patient/resident declined to respond [SKIP to E1d1]**
- ☐ 9 = **Unknown or unable to assess [SKIP to E1d1]**

E1c2. SYMPTOM FREQUENCY
ASK PATIENT/RESIDENT: "Over the last 2 weeks, how often have you been bothered by having 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)
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

E1d1. SYMPTOM PRESENCE
ASK PATIENT/RESIDENT: "Over the last 2 weeks, have you been bothered by feeling tired or having little energy?"

- ☐ 0 = No **[SKIP to E1e1]**
- ☐ 1 = Yes
- ☐ 7 = **Patient/resident declined to respond [SKIP to E1e1]**
- ☐ 9 = **Unknown or unable to assess [SKIP to E1e1]**

E1d2. SYMPTOM FREQUENCY
ASK PATIENT/RESIDENT: "Over the last 2 weeks, how often have you been bothered by feeling tired or having 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)
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

E1e1. SYMPTOM PRESENCE
ASK PATIENT/RESIDENT: "Over the last 2 weeks, have you been bothered by a poor appetite or overeating?"

- ☐ 0 = No **[SKIP TO E1f1]**
- ☐ 1 = Yes
- ☐ 7 = **Patient/resident declined to respond [SKIP TO E1f1]**
- ☐ 9 = **Unknown or unable to assess [SKIP TO E1f1]**

E1e2. SYMPTOM FREQUENCY
ASK PATIENT/RESIDENT: "Over the last 2 weeks, how often have you been bothered by 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)
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

E1f1. SYMPTOM PRESENCE
ASK PATIENT/RESIDENT: "Over the last 2 weeks, have you been bothered by feeling bad about yourself – or that you are a failure or have let yourself or your family down?"

- ☐ 0 = No **[SKIP TO E1g1]**
- ☐ 1 = Yes
- ☐ 7 = **Patient/resident declined to respond [SKIP TO E1g1]**
- ☐ 9 = **Unknown or unable to assess [SKIP TO E1g1]**

E1f2. SYMPTOM FREQUENCY
ASK PATIENT/RESIDENT: "Over the last 2 weeks, how often have you been bothered by feeling bad about yourself – or that you are a failure or have let yourself or your 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)
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

E1g1. SYMPTOM PRESENCE:
ASK PATIENT/RESIDENT: "Over the last 2 weeks, have you been bothered by trouble concentrating on things, such as reading the newspaper or watching television?"

- ☐ 0 = No **[SKIP TO E1h1]**
- ☐ 1 = Yes
- ☐ 7 = **Patient/resident declined to respond [SKIP TO E1h1]**
- ☐ 9 = **Unknown or unable to assess [SKIP TO E1h1]**

E1g2. SYMPTOM FREQUENCY
ASK PATIENT/RESIDENT: "Over the last 2 weeks, how often have you been bothered by 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)
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

E1h1. SYMPTOM PRESENCE
ASK PATIENT/RESIDENT: "Over the last 2 weeks, have you been bothered by moving or speaking so slowly that other people could have noticed. Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual?"

- ☐ 0 = No **[SKIP TO E1i1]**
- ☐ 1 = Yes
- ☐ 7 = **Patient/resident declined to respond [SKIP TO E1i1]**
- ☐ 9 = **Unknown or unable to assess [SKIP TO E1i1]**

E1h2. SYMPTOM FREQUENCY
ASK PATIENT/RESIDENT: "Over the last 2 weeks, how often have you been bothered by moving or speaking so slowly that other people could have noticed. Or the opposite - being so fidgety or restless that you have 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)
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

E1i1. SYMPTOM PRESENCE
ASK PATIENT/RESIDENT: "Over the last 2 weeks, have you been bothered by thoughts that you would be better off dead, or hurting yourself in some way?"

- ☐ 0 = No **[SKIP TO PHQ-9 TOTAL score]**
- ☐ 1 = Yes
- ☐ 7 = **Patient/resident declined to respond [SKIP TO PHQ-9 TOTAL score]**
- ☐ 9 = **Unknown or unable to assess [SKIP TO PHQ-9 TOTAL score]**

E1i2. SYMPTOM FREQUENCY
ASK PATIENT/RESIDENT: "Over the last 2 weeks, how often have you been bothered by thoughts that you would be better off dead, or hurting yourself in some way?"

- ☐ 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)
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

PHQ-9 TOTAL: **Add values from E1a2, E1b2, E1c2, E1d2, E1e2, E1f2, E1g2, E1h2, E1i2 and E1j2 →**

Testing Objectives

Basic descriptive statistics (e.g., frequencies, means, SDs) are presented for patient/resident responses to the PHQ-2 to 9 admission assessment administration. These descriptive statistics characterize the rates of depressive symptoms for patients/residents in each setting and for the overall sample. Data are presented for patients/residents who answered only the PHQ-2 and for those who continued and responded to all nine PHQ data elements. To examine known groups validity, we also examined PHQ-2 positive screen and PHQ-9 total scores by patient/resident characteristics and clinical groups of interest, both for the combined sample and for each setting separately. For admission data, feasibility (frequencies, rates of missingness, and time to complete) and interrater reliability (kappa and percent agreement) were examined. Lastly, frequencies at admission and discharge were compared to inform stability or possible change over time.

Results

Feasibility

Frequencies/Missing

Table 3.1 shows the percentage of responses at admission for PHQ-2 to 9 data elements, both overall and by setting. The PHQ-2 to 9 was administered to 3,010 of the 3,121 patients/residents, or 96 percent of the communicative admission sample: 639 in HHAs, 776 in IRFs, 479 in LTCHs, and 1,116 in SNFs. Among those who were administered the PHQ-2 to 9, missing data at the data element level ranged from 0.3 percent to 2.4 percent overall with minimal setting differences. Overall, results show that 28 percent of patients/residents screened positive on the PHQ-2 and proceeded to complete the full PHQ-9. Those who screened positive and completed

the full PHQ-9 are described in more detail below (in the "Known Groups Validity" section). At the setting level, positive screens ranged from 24 percent in HHAs to 38 percent in LTCHs. Eligibility for PHQ-9 per PHQ-2 (i.e., positive screen) was significantly higher for LTCH patients (38 percent; $\chi^2_{(3)} = 29.34$, $p < 0.0001$) relative to the other three settings, whose rates were similar (24 percent in HHAs, 27 percent in IRFs, and 27 percent in SNFs).

Among the 28 percent of patients/residents who screened positive on the PHQ-2 and completed the PHQ-9, the overall average score was 11.9 (SD = 5.3) out of a total possible score of 27 and ranged from 11.4 (SD = 5.0) in HHAs to 13.0 (SD = 5.8) in LTCHs. PHQ-9 summary scores were significantly associated with setting type ($F_{(3,842)} = 3.4$, $p = 0.02$), such that summary scores were significantly higher in LTCHs (mean [M] = 13.0, SD = 5.8) than in SNFs (M = 11.5, SD = 5.0), $t_{(842)} = 2.84$, $p < 0.01$. Overall, among patients/residents who completed the PHQ-9, the majority were classified as having either mild (31 percent) or moderate (32 percent) depression severity based on the Total Severity Score scale at the time of the assessment. LTCH patients, however, tended to be classified more often in the higher severity categories relative to patients/residents in the other settings ($\chi^2_{(12)} = 26.77$, $p < 0.01$).

Table 3.1. Overall and Setting-Specific Response Frequencies for PHQ-2 to 9 Data Elements at Admission (percent)

Data Element	HHA (*n* = 639)	IRF (*n* = 776)	LTCH (*n* = 479)	SNF (*n* = 1,116)	Overall (*n* = 3,010)
Symptom presence and frequency: little interest or pleasure (e1a)					
No	65	61	56	65	62
0–1 day	4	4	5	3	4
2–6 days	15	16	13	13	14
7–11 days (half or more)	9	10	11	9	10
12–14 days (nearly all)	8	10	16	10	11
Symptom presence and frequency: feeling down, depressed, hopeless (e1b)					
No	62	57	49	58	57
0–1 day	3	6	4	5	4
2–6 days	20	19	19	19	19
7–11 days (half or more)	7	9	13	8	9
12–14 days (nearly all)	8	8	16	11	10
PHQ-2					
Mean (SD)	2.2 (1.6)	2.3 (1.7)	2.7 (1.8)	2.4 (1.7)	2.4 (1.7)
Eligible for PHQ-9 per PHQ-2					
Yes	24	27	38	27	28
Number of patients/residents eligible for PHQ-9 per PHQ-2	153	209	182	306	850

Data Element	HHA (*n* = 639)	IRF (*n* = 776)	LTCH (*n* = 479)	SNF (*n* = 1,116)	Overall (*n* = 3,010)
Symptom presence and frequency: too little/too much sleep (e1c)					
No	30	34	34	33	33
0–1 day	2	3	1	2	2
2–6 days	15	15	13	16	15
7–11 days (half or more)	19	16	20	16	17
12–14 days (nearly all)	34	31	32	34	33
Symptom presence and frequency: tired/no energy (e1d)					
No	10	11	13	10	11
0–1 day	1	0	1	1	1
2–6 days	9	17	13	17	15
7–11 days (half or more)	27	26	23	28	26
12–14 days (nearly all)	52	46	50	44	48
Symptom presence and frequency: poor appetite or overeating (e1e)					
No	50	43	34	46	44
0–1 day	1	2	2	1	1
2–6 days	9	11	10	9	10
7–11 days (half or more)	17	13	16	15	15
12–14 days (nearly all)	22	31	39	29	30
Symptom presence and frequency: feel bad about self (e1f)					
No	55	52	51	58	55
0–1 day	1	2	1	1	1
2–6 days	12	12	12	10	12
7–11 days (half or more)	15	16	10	12	13
12–14 days (nearly all)	17	17	26	18	19
Symptom presence and frequency: trouble concentrating (e1g)					
No	54	47	44	48	48
0–1 day	1	1	1	1	1
2–6 days	15	16	9	16	14
7–11 days (half or more)	11	11	12	13	12
12–14 days (nearly all)	19	25	34	22	25
Symptom presence and frequency: moving or speaking slowly (e1h)					
No	64	62	50	68	62
0–1 day	1	0	2	1	1
2–6 days	9	9	10	7	9

Data Element	HHA (*n* = 639)	IRF (*n* = 776)	LTCH (*n* = 479)	SNF (*n* = 1,116)	Overall (*n* = 3,010)
7–11 days (half or more)	8	13	13	10	11
12–14 days (nearly all)	18	16	25	14	18
Symptom presence and frequency: suicidal thoughts (e1i)					
No	82	78	77	80	79
0–1 day	2	4	3	2	3
2–6 days	9	7	7	9	8
7–11 days (half or more)	5	3	5	5	4
12–14 days (nearly all)	3	7	7	4	5
PHQ-9					
Mean (SD)	11.4 (5.0)	11.8 (5.3)	13.0 (5.8)	11.5 (5.1)	11.9 (5.3)
Depression categorization (PHQ-9)					
Minimal (0–4)	10	4	6	7	6
Mild (5–9)	27	36	27	33	31
Moderate (10–14)	37	32	25	34	32
Moderately Severe (15–19)	20	19	28	18	21
Severe (20–27)	6	9	14	8	9

Known Groups Validity

Comparing the performance of patients/residents on the PHQ-2 to 9 data elements with other patient/resident characteristics adds information about the validity of data elements. If known or logical associations between patients/resident characteristics and data elements are observed in data from the National Beta test, this contributes to the evidence that the data elements are valid, or assessing the construct that they are intended to capture.

Table 3.2 shows the percentage of patients/residents with a positive PHQ-2 screen and the mean and SD of PHQ-9 total scores for the overall admission sample stratified by patient/resident characteristics and clinical groups as described in Chapter 1: gender (male or female as documented by National Beta Test assessor), age (18–44, 45–64, 65–74, 75–89, or 90 and over), length of stay (in days), disposition at discharge (e.g. to another PAC setting, home, hospital), sepsis, heart failure, stroke, and two ADLs: toileting (not available for HHA patients) and ability to transfer from lying to sitting. As a reminder, these clinical conditions were chosen based on their common occurrence across settings, their frequent relationship with many of the data elements tested in the National Beta Test, and their availability in all four settings (i.e., equivalent information was collected on the OASIS, IRF-PAI, LCDS, and MDS). Setting-specific results are presented in the appendix in Tables A.1–A.4.

Table 3.2. Frequencies for PHQ-2 Positive Screen and Mean (SD) for PHQ-9 Total Score by Patient/Resident Characteristics and Clinical Groups (percent)

Patient/Resident Characteristics and Clinical Groups	PHQ-2 Screen (% positive)	PHQ-9 Total Score M (SD)
Gender (n_{PHQ2} = 2,906[a]; n_{PHQ9} = 816)		
Male (n_{PHQ2} = 1,199; n_{PHQ9} = 312)	26.1	12.0 (5.6)
Female (n_{PHQ2} = 1,707; n_{PHQ9} = 504)	29.7	11.9 (5.2)
Age (n_{PHQ2} = 2,895[a]; n_{PHQ9} = 814[b])		
18–44 (n_{PHQ2} = 42; n_{PHQ9} = 16)	38.1	13.4 (5.8)
45–64 (n_{PHQ2} = 304; n_{PHQ9} = 122)	40.1	13.6 (5.8)
65–74 (n_{PHQ2} = 909; n_{PHQ9} = 271)	29.8	12.1 (5.5)
75–89 (n_{PHQ2} = 1,331; n_{PHQ9} = 351)	26.6	11.3 (5.0)
90+ (n_{PHQ2} = 309; n_{PHQ9} = 54)	17.8	10.9 (5.2)
Length of stay (n_{PHQ2} = 2,566; n_{PHQ9} = 721; mean, SD)	Pos: 21.6 (12.4) Neg: 21.5 (12.9)	Pearson r = 0.01
Disposition at discharge (n_{PHQ2} = 2,856[a]; n_{PHQ9} = 800)		
Home (n_{PHQ2} = 1,337; n_{PHQ9} = 327)	24.5	11.7 (5.1)
Hospital (n_{PHQ2} = 200; n_{PHQ9} = 67)	34.5	13.1 (5.8)
Hospice (n_{PHQ2} = 40; n_{PHQ9} = 9)	22.5	13.6 (5.0)
HHA (n_{PHQ2} = 621; n_{PHQ9} = 192)	31.1	11.7 (5.7)
IRF (n_{PHQ2} = 51; n_{PHQ9} = 20)	39.2	15.1 (6.2)
LTCH (n_{PHQ2} = 12; n_{PHQ9} = 2)	16.7	14.8 (0.3)
SNF (n_{PHQ2} = 279; n_{PHQ9} = 91)	32.6	11.4 (5.3)
Other (n_{PHQ2} = 316; n_{PHQ9} = 92)	29.1	11.9 (5.3)
Clinical conditions (n_{PHQ2} = 2,239; n_{PHQ9} = 635)		
Sepsis		
Yes (n_{PHQ2} = 151[a]; n_{PHQ9} = 55[b])	36.4	13.3 (5.7)
No (n_{PHQ2} = 2088; n_{PHQ9} = 580)	27.9	11.8 (5.3)
Heart failure		
Yes (n_{PHQ2} = 385; n_{PHQ9} = 106)	27.5	12.1 (5.1)
No (n_{PHQ2} = 1854; n_{PHQ9} = 529)	28.7	11.9 (5.4)
Stroke		
Yes (n_{PHQ2} = 192; n_{PHQ9} = 47)	24.5	12.0 (5.6)
No (n_{PHQ2} = 2047; n_{PHQ9} = 588)	28.9	12.0 (5.3)
Hygiene—Toileting (n_{PHQ2} = 1,510[a]; n_{PHQ9} = 453)		
Independent (n_{PHQ2} = 73; n_{PHQ9} = 22)	30.1	12.8 (6.2)
Setup or clean-up assistance (n_{PHQ2} = 77; n_{PHQ9} = 11)	14.3	11.1 (4.7)
Supervision or touching assistance (n_{PHQ2} = 322; n_{PHQ9} = 100)	31.1	11.6 (4.9)
Partial/moderate assistance (n_{PHQ2} = 359; n_{PHQ9} = 102)	28.7	11.6 (5.4)
Substantial/maximal assistance (n_{PHQ2} = 335; n_{PHQ9} = 94)	28.7	12.4 (5.1)

Patient/Resident Characteristics and Clinical Groups	PHQ-2 Screen (% positive)	PHQ-9 Total Score M (SD)
Dependent (n_{PHQ2} = 344; n_{PHQ9} = 123)	35.8	13.0 (5.8)
Mobility—Lying to sitting (n_{PHQ2} = 1,868[a]; n_{PHQ9} = 529)		
Independent (n_{PHQ2} = 192; n_{PHQ9} = 52)	27.1	12.4 (5.8)
Setup or clean-up assistance (n_{PHQ2} = 113; n_{PHQ9} = 25)	22.1	11.5 (5.5)
Supervision or touching assistance (n_{PHQ2} = 527; n_{PHQ9} = 140)	26.6	11.4 (4.7)
Partial/moderate assistance (n_{PHQ2} = 612; n_{PHQ9} = 169)	27.9	11.9 (51)
Substantial/maximal assistance (n_{PHQ2} = 291; n_{PHQ9} = 91)	31.3	12.1 (5.6)
Dependent (n_{PHQ2} = 133; n_{PHQ9} = 52)	39.1	13.2 (6.6)

[a] Significant ($p < 0.05$) associations with PHQ-2 positive screen as indicated by chi-square tests of independence.
[b] Significant ($p < 0.05$) mean differences in PHQ-9 total score as indicated by analysis of variance. Because of differences in sample sizes for PHQ-2 screening and PHQ-9 total scores, we report sample sizes for each (n_{PHQ2} = PHQ-2 screening; n_{PHQ9} = PHQ-9 total score).

Based on the research literature, we generated several hypotheses or expectations for associations between the data elements and the patient/resident characteristics. We expected signs and symptoms of depression as indicated by patient/resident response to the PHQ-2 to 9 to be related to gender, age, stroke, toileting, and mobility (ability to transfer from lying to sitting), such that patients/residents with a positive screen for signs and symptoms of depression would tend to be female,[68] younger,[69] to have had a stroke,[70] and to have less independence in ADLs.[71]

Overall, both positive screen on the PHQ-2 and total PHQ-9 scores were significantly associated with age. Additionally, PHQ-2 positive screen was associated with gender, disposition at discharge, sepsis, toileting, and ability to transfer from lying to sitting. There were no other significant associations between PHQ-2 positive screen and PHQ-9 summary score with any of the remaining patient/resident characteristics or clinical groups. We review the statistical associations between variables and give more details on the directions of the associations in the bullets below.

Gender and Age

- Gender, overall, was significantly associated with a positive screen on the PHQ-2 ($\chi^2_{(1)} = 4.5$, $p < 0.05$), such that a greater percentage of females (29.7 percent) screened positive compared with males (26.1 percent). Similar trends were observed at the setting level in HHAs and SNFs but not in IRFs or LTCHs. Gender was not associated with PHQ-9 summary scores overall or in any of the settings. The lack of associations within some settings and differences in patterns of association between PHQ-2 and PHQ-9 scores is likely due to the sample size. For example, fewer patients/residents completed the PHQ-9

[68] Djernes 2006; Jones, Marcantonio, and Rabinowitz, 2003.

[69] Jones, Marcantonio, and Rabinowitz, 2003.

[70] Jongenelis et al., 2004.

[71] Kaup et al., 2007.

than the PHQ-2, and similar differences that would attain statistical significance in a larger sample are less likely to reach significance in a smaller sample. *These findings are consistent with our expectation to see higher rates of the signs and symptoms of depression in women than men, although statistically significant differences were only observed in the PHQ-2 scores and not in all settings.*

- In the overall sample, age was significantly associated with both PHQ-2 positive screen ($\chi^2_{(4)} = 42.72$, $p < 0.01$) and PHQ-9 total score ($F_{(4,809)} = 5.04$, $p < 0.01$), with a general trend of fewer depressive symptoms with increasing age. For the PHQ-2 positive screen, the 45–64-year-old age group had the highest positive screen rate (40 percent) compared with the 90+ (18 percent) and 75–89 (27 percent) age groups. At the setting level, age was significantly associated with PHQ-2 positive screen in HHAs and SNFs but not in IRFs or LTCHs. Specifically, in both HHAs and SNFs, the 90+ age group had the lowest positive screen rate, 19 percent and 18 percent, respectively. As in the overall sample, in HHAs the greatest positive screen rate was for patients 45–64 years of age (40 percent). In SNFs the highest positive screen rate was for residents 18–44 years of age (56 percent). For patients/residents who screened positive on the PHQ-2 and continued on to the PHQ-9, patients/residents in the 45–64 age group had the highest PHQ-9 summary score (M = 13.6, SD = 5.8), which was significantly higher than the 90+ age group (M = 10.9, SD = 5.2, $t_{(809)} = 3.18$, $p < 0.01$). A similar significant association between age and PHQ-9 summary score at the setting level was found only in SNFs. Age was not associated with PHQ-9 summary scores in any of the remaining settings, due either to the lower sample size of patients completing the full PHQ-9 in IRFs, LTCHs, and HHAs or because this relationship of age to PHQ-9 does not hold true in all PAC settings. *These findings are consistent with our expectation to see lower rates of the signs and symptoms of depression in older patients/residents, although statistically significant differences observed in the overall sample were noted only in some settings.*

Length of Stay, Disposition at Discharge

- Length of stay, overall, was not significantly associated with screening positive on the PHQ-2, such that length of stay for those who screened positive (M = 21.6 days, SD = 12.4) was not different from those who screened negative (M = 21.5 days, SD = 12.9). Similarly, length of stay was not associated with PHQ-9 summary scores overall ($r = 0.01$). At the setting level, length of stay was not significantly associated with either positive screen for PHQ-2 or PHQ-9 summary scores in any of the settings. *We did not have any expectations about the relationship between PHQ-2 and PHQ-9 results and length of stay.*
- Overall, disposition at discharge was significantly associated with positive screen on PHQ-2, ($\chi^2_{(7)} = 22.67$, $p < 0.001$). Relative to patients/residents discharged to all other placements (e.g., SNF, LTCH, home, hospital), a greater percentage of those discharged to IRFs (39 percent) had screened positive on the PHQ-2. At the setting level, disposition at discharge was significantly associated with positive screen on PHQ-2 only in IRFs. When looking at where IRF patients in our sample were discharged, relative to patients/residents discharged to all other settings, a greater percentage of patients discharged to hospitals (42 percent) had screened positive on the PHQ-2. Disposition at discharge was not significantly associated with PHQ-9 summary scores overall or in any of the settings. *We did not have any expectations about the relationship between PHQ-2*

and PHQ-9 results and disposition at discharge. However, this pattern of findings suggests that patients with ongoing rehabilitation needs (e.g., those being discharged to an IRF) or IRF patients who are discharged to a hospital may be especially likely to exhibit signs and symptoms of depression. This could reflect underlying medical conditions in these patients that are known to correlate with depression, such as stroke, as noted above.

Clinical Conditions

- There was an overall significant association of sepsis with positive screen on PHQ-2 ($\chi^2_{(1)} = 5.00$, $p < 0.05$), such that a greater percentage of patients/residents with sepsis screened positive on PHQ-2 (36.4 percent) compared with those without (27.9 percent). A statistically significant association was not observed in any of the settings. *We did not have any expectations about the relationship between PHQ-2 and sepsis. However, in the overall sample, the finding of more depressive symptoms among the most seriously ill patients/residents is not surprising.*
- There were no associations of heart failure with positive screen on PHQ-2 overall or in any of the settings. *We did not have any expectations about the relationship between PHQ-2 and heart failure.*
- There were no associations of stroke with positive screen on PHQ-2 overall or in any of the settings. *Based on prior studies referenced above, we expected to find an association between stroke and the PHQ-2, but it was not present in our data, perhaps because of the relatively small proportion of patients with stroke in this sample.*

ADLs: Toileting and Ability to Transfer from Lying to Sitting

- In the overall sample, PHQ-2 positive screen was associated with independence levels on both toileting ($\chi^2_{(5)} = 15.18$, $p < 0.01$) and ability to transfer from lying to sitting ($\chi^2_{(5)} = 11.95$, $p < 0.05$), such that those patients/residents who were completely dependent for these ADLs were more likely to screen positive on the PHQ-2. This trend was also observed among SNF residents for toileting ($\chi^2_{(5)} = 11.50$, $p < 0.05$) and among LTCH patients for both toileting ($\chi^2_{(5)} = 14.70$, $p < 0.05$) and ability to transfer from lying to sitting ($\chi^2_{(5)} = 13.80$, $p < 0.05$). PHQ-9 summary scores were not associated with ADLs in the overall sample or in any of the settings. *These findings generally conform to our expectations that patients/residents with more ADL independence would display fewer signs and symptoms of depression.*

Time to Complete

Table 3.3 shows the average time to complete the PHQ-2 to 9 overall and by setting. Time to complete is broken down into three categories: "All" (everyone regardless of whether they screened negative on PHQ-2), "PHQ-2 Only" (those who completed the PHQ-2, screened negative, and skipped the remaining data elements), and "PHQ-2 to 9" (those who screened positive on PHQ-2 and continued on to the remaining PHQ-9 data elements).

On average, the time to complete for "All" was 2.3 minutes (SD = 1.5) and ranged from 2.0 minutes (SD = 1.3) in IRFs to 2.6 minutes (SD = 1.6) in LTCHs. Time to complete was significantly associated with setting type ($F_{(3,1723)} = 12.63$, $p < 0.01$), such that it took

significantly more time to complete in LTCHs than in the other three settings (IRFs [$t_{(1723)} = 6.1$, $p < 0.01$), HHAs [$t_{(1723)} = 3.4$, $p < 0.01$, and SNFs [$t_{(1723)} = 3.15$, $p < 0.01$]). Further, it took less time to complete in IRFs than in HHAs ($t_{(1723)} = 2.84$, $p < 0.01$) and SNFs ($t_{(1723)} = 3.31$, $p < 0.01$). HHAs did not significantly differ from SNFs.

For those completing only the PHQ-2, on average, time to complete was 1.7 minutes (SD = 1.1) and ranged from 1.5 minutes (SD = 0.9) in IRFs to 1.9 minutes (SD = 1.3) in LTCHs. Time to complete was significantly associated with setting type ($F_{(3,1298)} = 8.07$, $p < 0.01$), such that it took significantly less time to complete in IRFs than in LTCHs ($t_{(1298)} = 4.5$, $p < 0.01$) and HHAs ($t_{(1298)} = 3.57$, $p < 0.01$). There were no other differences among settings.

On average, those completing the full PHQ-2 to 9 took 4.0 minutes (SD = 1.2), ranging from 3.7 minutes (SD = 1.2) in IRFs to 4.2 minutes (SD = 1.2) in HHAs. Time to complete was significantly associated with setting type ($F_{(3,410)} = 3.49$, $p < 0.05$); however, after corrections for multiple comparisons, no statistically significant differences among settings emerged. That said, time to complete tended to be shorter in IRFs than in other settings.

Table 3.3. Time to Complete the PHQ-2 to 9 (minutes)

	HHA (*n* = 428)	IRF (*n* = 515)	LTCH (*n* = 305)	SNF (*n* = 479)	Overall (*n* = 1,727)
All					
Mean (SD)	2.3 (1.5)	2.0 (1.3)	2.6 (1.6)	2.3 (1.5)	2.3 (1.5)
PHQ-2 Only					
Mean (SD)	1.8 (1.2)	1.5 (0.9)	1.9 (1.3)	1.7 (1.1)	1.7 (1.1)
PHQ-2 to 9					
Mean (SD)	4.2 (1.2)	3.7 (1.2)	4.1 (1.3)	4.1 (1.2)	4.0 (1.2)

Time to complete was also evaluated according to urbanicity (urban versus nonurban), geographic region (Northeast, South, Midwest, West), facility ownership (for-profit versus nonprofit) and facility size (above or below setting-type median) to evaluate the generalizability of these performance results (see Tables A.5–A.8 in the appendix). No significant differences were found for time to complete the PHQ-2 to 9 in these sensitivity analyses.

Interrater Reliability

Table 3.4 shows the kappa interrater reliability coefficients for the PHQ-2 to 9 overall and by setting. As described in more detail in Volume 3, paired assessment data for interrater reliability evaluation was collected on a subset of the National Beta Test communicative admission sample of patients/residents according to setting-level target totals. For example, each participating LTCH was asked to conduct 20 paired assessments to contribute to interrater reliability. Inclusion in interrater reliability data collection depended on paired facility staff and research nurse assessors' ability to schedule assessments. Kappas were computed on 948

patients/residents who were assessed by research nurse and facility/agency staff assessor pairs: 196 in HHAs, 254 in IRFs, 231 in LTCHs, and 267 in SNFs. Overall kappas for the PHQ-2 to 9 were excellent, ranging from 0.95 to 1.00. At the setting level, kappas were also excellent, ranging from 0.92 to 1.00 in HHAs, 0.87 to 1.00 in IRFs, 0.94 to 1.00 in LTCHs, and 0.94 to 1.00 in SNFs.

Table 3.4. Interrater Reliability Kappa or Weighted Kappa for PHQ Data Elements

Data Element	HHA (*n* = 196)	IRF (*n* = 254)	LTCH (*n* = 231)	SNF (*n* = 267)	Overall (*n* = 948)
Symptom present: little interest or pleasure (e1a1)	0.95	0.99	0.99	0.98	0.98
Symptom frequency: little interest or pleasure (e1a2)	0.98	1.00	0.98	0.98	0.99
Symptom present: feeling down, depressed, hopeless (e1b1)	0.99	0.98	1.00	0.99	0.99
Symptom frequency: feeling down, depressed, hopeless (e1b2)	0.93	0.98	0.98	0.99	0.98
Eligible for PHQ-9 per PHQ-2	0.96	0.98	0.98	0.98	0.98
Symptom present: too little/too much sleep (e1c1)	0.90	1.00	1.00	1.00	0.98
Symptom frequency: too little/too much sleep (e1c2)	1.00	0.98	0.90	0.96	0.96
Symptom present: tired/no energy (e1d1)	1.00	0.91	0.95	0.94	0.95
Symptom frequency: tired/no energy (e1d2)	1.00	0.93	0.98	1.00	0.98
Symptom present: poor appetite or overeating (e1e1)	0.96	0.93	0.95	1.00	0.96
Symptom frequency: poor appetite or overeating (e1e2)	1.00	1.00	1.00	1.00	1.00
Symptom present: feel bad about self (e1f1)	1.00	1.00	1.00	1.00	1.00
Symptom frequency: feel bad about self (e1f2)	1.00	1.00	0.95	1.00	0.98
Symptom present: trouble concentrating (e1g1)	1.00	1.00	1.00	0.97	0.99
Symptom frequency: trouble concentrating (e1g2)	0.96	0.97	0.94	1.00	0.97
Symptom present: moving or speaking slowly (e1h1)	1.00	0.94	0.90	1.00	0.95
Symptom frequency: moving or speaking slowly (e1h2)	1.00	0.87	1.00	1.00	0.97
Symptom present: suicidal thoughts (e1i1)	1.00	1.00	0.94	1.00	0.98
Symptom frequency: suicidal thoughts (e1i2)	0.93	1.00	0.95	1.00	0.97
Sum of all symptom frequencies (PHQ-9)[a]	0.92	1.00	0.94	0.97	0.96

[a] As classified into the five categories shown in Table A.1. Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

Interrater reliability (kappa) was also evaluated according to urbanicity (urban versus nonurban), geographic region (Northeast, South, Midwest, West), facility ownership (for-profit versus nonprofit), and facility size (above or below setting-type median) to evaluate the generalizability of these performance results (see Tables A.9–A.12 in the appendix). No differences were found for interrater reliability of the PHQ-2 to 9 in these sensitivity analyses.

Table 3.5 shows the percent agreement for the PHQ-2 to 9 overall and by setting. Overall, percent agreement was high for all data elements, ranging from 97 to 100 percent with minimal

setting differences. At the setting level, percent agreement ranged from 93 to 100 percent in HHAs, 93 to 100 percent in IRFs, 94 to 100 percent in LTCHs, and 96 to 100 percent in SNFs.

Table 3.5. Interrater Reliability—Percent Agreement for PHQ Data Elements

Data Element	HHA (*n* = 196)	IRF (*n* = 254)	LTCH (*n* = 231)	SNF (*n* = 267)	Overall (*n* = 948)
Symptom present: little interest or pleasure (e1a1)	97	100	100	99	99
Symptom frequency: little interest or pleasure (e1a2)	99	100	98	98	99
Symptom present: feeling down, depressed, hopeless (e1b1)	99	99	100	100	100
Symptom frequency: feeling down, depressed, hopeless (e1b2)	95	98	98	99	98
Eligible for PHQ-9 per PHQ-2	98	99	99	99	99
Symptom present: too little/too much sleep (e1c1)	96	100	100	100	99
Symptom frequency: too little/too much sleep (e1c2)	100	98	94	96	97
Symptom present: tired/no energy (e1d1)	100	98	99	99	99
Symptom frequency: tired/no energy (e1d2)	100	96	99	100	99
Symptom present: poor appetite or overeating (e1e1)	98	97	97	100	98
Symptom frequency: poor appetite or overeating (e1e2)	100	100	100	100	100
Symptom present: feel bad about self (e1f1)	100	100	100	100	100
Symptom frequency: feel bad about self (e1f2)	100	100	95	100	98
Symptom present: trouble concentrating (e1g1)	100	100	100	99	100
Symptom frequency: trouble concentrating (e1g2)	96	97	97	100	98
Symptom present: moving or speaking slowly (e1h1)	100	97	95	100	98
Symptom frequency: moving or speaking slowly (e1h2)	100	93	100	100	98
Symptom present: suicidal thoughts (e1i1)	100	100	98	100	99
Symptom frequency: suicidal thoughts (e1i2)	93	100	95	100	97
Sum of all symptom frequencies (PHQ-9)[a]	94	100	95	97	97

[a] As classified into the five categories shown in Table 3.1.

Admission to Discharge

Table 3.6 summarizes patterns of change on PHQ-2 to 9 data elements from admission to discharge. As described in more detail in Volume 3, discharge data were collected on a subset of the National Beta Test admission sample of patients/residents. Availability of discharge data depended on advance notification of discharge and the ability to schedule assessments among the facility staff assessors in each participating site. Patterns are characterized as "no change" (scores stay the same at admission and discharge), "more severe" (scores indicate a decline from admission to discharge), and "less severe" (scores improve from admission to discharge). For the PHQ-2 to 9, admission and discharge data were collected on 781 patients/residents: 144 in HHAs, 335 in IRFs, 81 in LTCHs, and 221 in SNFs. Overall, while responses to the PHQ-2 demonstrated some degree of stability from admission to discharge, compared with admission,

patients/residents were less symptomatic at discharge: They were less likely to have little interest or pleasure ($t_{(755)} = 8.72$, $p < 0.001$) or feel down, depressed, or hopeless ($t_{(776)} = 9.23$, $p < 0.001$). Compared with admission, patients/residents at discharge were less likely to be eligible for PHQ-9 per PHQ-2 ($t_{(780)} = 8.27$, $p < 0.001$). Moreover, for patients/residents who screened positive on PHQ-2 at both admission and discharge, scores tended to reflect improvement in depressive symptoms from admission to discharge, such that the distribution of scores into the PHQ-9 severity categories was significantly different ($t_{(779)} = 8.8$, $p < 0.001$). That is, of the 48 patients/residents who completed the PHQ-9 at both admission and discharge, scores for 35 percent placed them in less severe PHQ-9 depression categories at discharge relative to admission.

Table 3.6. Admission to Discharge Results for PHQ-2 and PHQ-9 Data Elements (percent)

Data Element	HHA (*n* = 144)	IRF (*n* = 335)	LTCH (*n* = 81)	SNF (*n* = 221)	Overall (*n* = 781)
Symptom presence and frequency: little interest or pleasure (e1a)					
No change	67	60	53	59	60
More severe	6	11	9	12	10
Less severe	28	29	38	29	30
Symptom presence and frequency: feeling down, depressed, hopeless (e1b)					
No change	73	62	55	65	64
More severe	3	9	10	8	8
Less severe	24	28	35	27	28
Eligible for PHQ-9 per PHQ-2					
No change	80	78	67	76	77
More severe	1	5	9	5	4
Less severe	19	17	25	19	19
Number of patients/residents eligible for PHQ-9 at both times	3	29	6	10	48
Depression categorization (PHQ-9)					
No change	67	59	67	70	63
More severe	0	3	0	0	2
Less severe	33	38	33	30	35

Assessor Feedback

Standardizing the assessment of depressive symptoms was important to assessors. They thought the PHQ was clinically relevant and provided information that was important to convey when patients/residents are transitioning between settings. On average, when asked to rate the clinical usefulness of the PHQ, facility/agency staff and research nurse assessors rated the PHQ

in the middle to high range on clinical utility in the assessor survey, relative to other data elements. When discussing the PHQ in focus groups, one facility staff member suggested that depression should be included on a standardized assessment and transfer form:

> If you have somebody that's very, very nervous or very, very depressed, they're not going to progress along as somebody that doesn't.
>
> —Boston, SNF Staff

Although there was general appreciation for the range of questions asked in the PHQ, facility staff focus group participants felt that neither PHQ option in the National Beta Test struck the right balance between depth and brevity (i.e., PHQ-2 was not enough, PHQ-9 was too much).

Some staff thought certain questions on the PHQ had the potential to cause patients distress, which could disrupt workflow and possibly complicate treatment:

> For the IRF especially, . . . we get a lot of traumatic accidents—spinal cords, traumatic brain injuries, things like that—and I had several times that we got to the questions that were like, "Do you feel hopeless?" and literally their face drained and they went, "Are you really asking me that question right now? I am never going to walk again. My husband just died three days ago in an accident." And they just shut down. That's it. You can't get past it at that point. And that happened more than I thought it would.
>
> —Phoenix, IRF Staff

Facility staff disliked the question "Over the past two weeks, have you been bothered by feeling down, depressed, or hopeless?" as it often invoked negative emotional responses among patients/residents. Additionally, staff noted that some patients/residents found use of these three distinct words (*down*, *depressed*, and *hopeless*) together to indicate the same concept to be confusing, and there was uncertainty about whether patients were being forthcoming with their answers.

A challenge to administration noted in the research nurses' focus groups was the two-week recollection time period required by the PHQ. First, patients/residents had difficulty recalling emotion during this long of a time period. Second, depending on the PAC setting, this period frequently included an acute care stay. Patients/residents would report that their mental status had changed with the change in circumstances, which could not be captured by these data elements. Moreover, research nurses believed that in combination with some of the longer questions on the PHQ, the two-week period was difficult to answer for the sickest patients, particularly those with impaired concentration and memory. Facility staff similarly reported the perception that the PHQ was burdensome for some patients to answer.

In summary, assessors in focus groups had some concerns with the burden of collection but generally appreciated the clinical utility and relevance of the PHQ, noting the importance of standardizing the assessment of depressive symptoms. As one research nurse articulated, "Just because it is difficult doesn't mean you shouldn't ask."

Summary

Results for the PHQ-2 to 9 data elements indicate moderate overall support for cross-setting standardization. Assessors emphasized the importance and high clinical utility of assessing mood in their patients, despite the personal and potentially sensitive nature of some of the data elements. Assessor feedback indicates that successful implementation and use of the PHQ in practice will require training and familiarity. The use of the PHQ-2 as a screen for the full PHQ-9 appears to mitigate the burden of the PHQ administration considerably. Nearly 75 percent of patients/residents overall screened negative on the PHQ-2 and completed the mood assessment in less than two minutes on average. Observed associations of the PHQ-2 screen and PHQ-9 total score with patient characteristics provide support for the validity of the mood assessment. For example, patients/residents completely dependent in mobility and toileting ADLs were more likely to screen positive on the PHQ-2. In terms of psychometric performance, kappas for the PHQ-2 to 9 were excellent and percent agreement was high for all data elements, with minimal setting differences. Responses to the PHQ-2 demonstrated some degree of stability from admission to discharge, but patients/residents were less symptomatic at discharge. These combined results show high feasibility, excellent interrater reliability, and high clinical utility for the PHQ-2 to 9 as a candidate data element for standardization across PAC settings.

4. PROMIS Depression

Data Element Description

The PROMIS Depression data elements include eight data elements from the PROMIS Depression Item Bank that assess depression across a wide variety of symptoms, covering negative mood (sadness, guilt), views of self (self-criticism, worthlessness), and social cognition (loneliness, interpersonal alienation), as well as decreased positive affect and engagement (loss of interest, meaning, and purpose). Please see Chapter 2 in this volume for more information on this data element selection process.

The PROMIS Depression data elements are completed through patient/resident interview. These data elements are not currently included in any PAC assessment instruments. In the National Beta Test, the PROMIS Depression data elements were collected in two versions. One version was identical to the format in the PROMIS Depression Item Bank, using a reference to "the past 7 days" for symptoms. The second version, shown in Figure 4.1, was identical except that it referenced "the past 3 days" to align more closely with PAC patient/resident assessment data elements that are currently in use.

Figure 4.1. PROMIS Depression Data Elements

SAY TO PATIENT/RESIDENT: "I am now going to ask you about your emotional distress, specifically depression and how you have been feeling over the past 3 days." I will also ask about some common problems that sometimes go along with feeling depressed. This is not meant to give you a diagnosis. Some of the questions might seem personal, but all patients/residents are asked to answer them. Knowing the answers to these questions will help us provide you with a more individualized care plan."

E2a. ASK PATIENT/RESIDENT:
"In the past 3 days, I felt worthless:"

- ☐ 1 = Never
- ☐ 2 = Rarely
- ☐ 3 = Sometimes
- ☐ 4 = Often
- ☐ 5 = Always
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

E2b. ASK PATIENT/RESIDENT:
"In the past 3 days, I felt that I had nothing to look forward to:"

- ☐ 1 = Never
- ☐ 2 = Rarely
- ☐ 3 = Sometimes
- ☐ 4 = Often
- ☐ 5 = Always
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

E2c. ASK PATIENT/RESIDENT:
"In the past 3 days, I felt helpless:"

- ☐ 1 = Never
- ☐ 2 = Rarely
- ☐ 3 = Sometimes
- ☐ 4 = Often
- ☐ 5 = Always
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

E2d. ASK PATIENT/RESIDENT:
"In the past 3 days, I felt sad:"

- ☐ 1 = Never
- ☐ 2 = Rarely
- ☐ 3 = Sometimes
- ☐ 4 = Often
- ☐ 5 = Always
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

E2e. ASK PATIENT/RESIDENT:
"In the past 3 days, I felt lonely:"

- ☐ 1 = Never
- ☐ 2 = Rarely
- ☐ 3 = Sometimes
- ☐ 4 = Often
- ☐ 5 = Always
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

E2f. ASK PATIENT/RESIDENT:
"In the past 3 days, I felt depressed:"

- ☐ 1 = Never
- ☐ 2 = Rarely
- ☐ 3 = Sometimes
- ☐ 4 = Often
- ☐ 5 = Always
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

E2g. ASK PATIENT/RESIDENT:
"In the past 3 days, I felt I had no reason for living:"

- ☐ 1 = Never
- ☐ 2 = Rarely
- ☐ 3 = Sometimes
- ☐ 4 = Often
- ☐ 5 = Always
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

E2h. ASK PATIENT/RESIDENT:
"In the past 3 days, I felt hopeless:"

- ☐ 1 = Never
- ☐ 2 = Rarely
- ☐ 3 = Sometimes
- ☐ 4 = Often
- ☐ 5 = Always
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

Testing Objectives

As stated above, two versions of the PROMIS Depression data elements were administered during the National Beta Test: a standard (past 7 days) version and a modified (past 3 days) version. The 7-day version was administered in Market Group A, and the 3-day version in Market Group B. For more details on Market Group A and B samples and characteristics, please refer to Volume 3. Below, we first present basic descriptive statistics (e.g., frequencies and scale scores) for admission data to characterize the responses to the PROMIS Depression data elements according to version for patients/residents in each setting and for the overall sample.

Next, we present differences in the versions that we evaluated by (1) conducting t-tests to compare data element and scale level means across the two versions (a Cohen's *d* value greater than 0.2 was used to indicate meaningful version differences) and (2) conducting DIF by version to determine whether data elements functioned differently within the scale according to version.

Because we saw little evidence of performance differences according to version in our analyses of the National Beta Test data (described further below), the majority of results, with the exception of feasibility results (data element frequencies, rates of missingness, and time to complete), are presented for the two versions combined. Specifically, we used the combined admission response data to (1) evaluate DIF according to setting, gender, and age, (2) examine PROMIS Depression data element T-scores by patient/resident characteristics and clinical groups of interest (i.e., known groups validity), and (3) evaluate interrater reliability (kappa and percent agreement), data element–total correlations, Cronbach's alpha internal consistency values, and unidimensionality indices. Also, frequencies at admission and discharge were compared for the combined versions to inform stability or possible change over time. Finally, because this set of data elements comes from the IRT-calibrated PROMIS Depression Item Bank, we provide a summed score to T-score conversion table in the appendix. This allows for the calculation of T-scores with M = 50 and SD = 10 corresponding to the total score on the eight National Beta Test data elements tested, allowing results to be directly compared with general population PROMIS Depression values.

Results

Feasibility

Frequencies/Missing

Tables 4.1 and 4.2 show the percentage of responses for each PROMIS Depression data element, both overall and by setting. Table 4.1 shows the results for Market Group A, which was administered version A asking about symptoms in the past 7 days, and Table 4.2 shows the results for Market Group B, which was administered version B asking about symptoms in the past 3 days. Version A was administered to 1,498 patients/residents: 228 in HHAs, 485 in IRFs, 249 in LTCHs, and 536 in SNFs. Version B was administered to 1,488 patients/residents: 409 in HHAs, 287 in IRFs, 224 in LTCHs, and 568 in SNFs. Overall, a little over 96 percent of the sample was administered one of the two versions. Among these patients/residents, overall missing data at the data element level ranged from 0.80 percent to 2.14 percent in version A (0.00 to 2.19 percent in HHAs, 0.21 to 1.86 percent in IRFs, 0.80 to 2.81 percent in LTCHs, and 0.56 to 2.99 percent in SNFs) and 0.81 percent to 1.48 percent in version B (0.24 to 1.22 percent in HHAs, 0.00 to 0.35 percent in IRFs, 0.89 to 2.68 percent in LTCHs, and 0.70 to 2.29 percent in SNFs) with minimal setting differences.

The tables present percentages of patients/residents endorsing frequencies of symptoms from "never" to "always." Across the two versions, there are minimal differences. For all settings and all data elements, the most frequently endorsed response category is "never." Across the settings, however, LTCH patients tend to endorse the highest frequency category (i.e., "always") consistently more than those from the other settings. For example, 6 percent of LTCH respondents in Market Group A endorsed the "always" category in response to the "worthless" item, whereas only 1 percent from HHAs, 2 percent from IRFs, and 2 percent from SNFs endorsed this category. This aligns with the overall depression data element scores being slightly higher in LTCHs relative to the other three settings. The tables also show the mean of the total depression score (range from 0–40) and the T-score (a standardized metric normed to the general population with M = 50 and SD = 10), overall and by setting across two different versions. The same pattern is observed for the total and the T-scores as in individual items. That is, there is little difference between the two versions, but, similar to PHQ-2 to 9 results, LTCH patients consistently have higher depression scores (total scores of 16.6 and 16.9 for versions A and B, respectively; T-scores of 53.7 and 53.9 for versions A and B, respectively) than patients/residents from the other settings (total scores averaging approximately 14 and T-scores averaging approximately 50). Formal analyses to evaluate differences in versions are presented below.

Table 4.1. Overall and Setting-Specific Response Frequencies for PROMIS Depression Data Elements at Admission: Market Group A (percent)

Data Element	HHA (*n* = 228)	IRF (*n* = 485)	LTCH (*n* = 249)	SNF (*n* = 536)	Overall (*n* = 1,498)
How often felt: worthless (e2a)					
Never	68	60	57	70	64
Rarely	15	15	10	8	12
Sometimes	11	18	21	15	16
Often	5	5	5	4	5
Always	1	2	6	2	2
How often felt: nothing to look forward to (e2b)					
Never	68	66	56	70	66
Rarely	14	14	14	10	13
Sometimes	14	13	18	12	13
Often	4	6	7	6	6
Always	0	1	5	2	2
How often felt: helpless (e2c)					
Never	55	42	38	48	45
Rarely	18	14	14	12	14

Sometimes	17	26	27	25	25
Often	7	13	11	11	11
Always	4	6	9	4	5
How often felt: sad (e2d)					
Never	44	33	28	41	37
Rarely	20	22	15	20	20
Sometimes	24	29	36	27	29
Often	7	12	14	9	11
Always	4	3	7	4	4
How often felt: lonely (e2e)					
Never	58	50	40	55	51
Rarely	15	19	15	12	15
Sometimes	19	18	26	22	21
Often	5	9	11	7	8
Always	4	4	9	4	5
How often felt: depressed (e2f)					
Never	55	48	43	57	51
Rarely	17	20	17	13	16
Sometimes	18	20	26	20	21
Often	8	8	7	6	7
Always	3	4	8	4	4
How often felt: no reason for living (e2g)					
Never	83	82	76	86	83
Rarely	7	8	8	5	7
Sometimes	9	6	10	5	7
Often	1	3	3	3	3
Always	0	1	3	1	1
How often felt: hopeless (e2h)					
Never	74	69	64	73	70
Rarely	12	13	13	8	11
Sometimes	10	12	14	11	12
Often	3	4	5	5	4
Always	1	2	4	2	2
Depression total Mean (SD)	13.5 (6.4)	14.9 (6.8)	16.6 (7.6)	14.2 (6.9)	14.7 (7.0)
Depression T-score Mean (SD)	49.2 (9.3)	51.6 (9.3)	53.7 (9.9)	50.2 (9.8)	51.1 (9.7)

Table 4.2. Overall and Setting-Specific Response Frequencies for PROMIS Depression Data Elements at Admission: Market Group B (percent)

Data Element	HHA (*n* = 409)	IRF (*n* = 287)	LTCH (*n* = 224)	SNF (*n* = 568)	Overall (*n* = 1,488)
How often felt: worthless (e2a)					
Never	64	63	53	59	60
Rarely	13	15	17	13	14
Sometimes	17	14	19	18	17
Often	6	5	6	6	6
Always	1	3	5	3	3
How often felt: nothing to look forward to (e2b)					
Never	67	62	55	65	63
Rarely	15	15	15	14	14
Sometimes	14	15	17	15	15
Often	4	5	7	5	5
Always	1	2	5	2	2
How often felt: helpless (e2c)					
Never	53	45	42	48	48
Rarely	16	16	15	13	15
Sometimes	21	23	24	25	23
Often	7	10	12	8	9
Always	2	5	8	6	5
How often felt: sad (e2d)					
Never	43	42	28	39	39
Rarely	20	21	19	18	19
Sometimes	26	24	30	30	28
Often	8	8	15	8	9
Always	3	6	7	5	5
How often felt: lonely (e2e)					
Never	58	54	34	48	50
Rarely	12	15	18	16	15
Sometimes	19	19	25	25	22
Often	8	9	12	7	8
Always	2	3	10	5	5
How often felt: depressed (e2f)					
Never	54	50	37	51	49
Rarely	18	21	16	16	18

Data Element	HHA (*n* = 409)	IRF (*n* = 287)	LTCH (*n* = 224)	SNF (*n* = 568)	Overall (*n* = 1,488)
Sometimes	19	19	27	21	21
Often	6	6	13	8	8
Always	3	5	6	4	4
How often felt: no reason for living (e2g)					
Never	84	83	72	81	81
Rarely	7	7	11	8	8
Sometimes	6	7	11	6	7
Often	2	2	4	3	3
Always	0	1	3	2	2
How often felt: hopeless (e2h)					
Never	75	73	61	72	71
Rarely	11	11	16	10	11
Sometimes	9	10	13	12	11
Often	4	4	6	4	4
Always	1	2	5	2	2
Depression total Mean (SD)	13.6 (6.4)	14.4 (6.8)	16.9 (8.2)	14.9 (7.2)	14.8 (7.1)
Depression T-score Mean (SD)	49.7 (9.2)	50.6 (9.5)	53.9 (10.5)	51.3 (9.7)	51.1 (9.7)

Version Comparison and Differential Item Functioning

Results of the comparison of data element and scale means by version are shown in Table A.13 in the appendix. While one data element ("I felt hopeless") out of eight showed statistically significant differences by version, this and all other differences were below the Cohen's *d* value of 0.2. At the scale level, differences were not significant and Cohen's *d* values were well below 0.2. In addition, results of evaluation of DIF according to version (Table A.14 in the appendix) showed that all data elements had R^2 effect sizes well below the 0.02 cut-off for nontrivial DIF. This means that the data element properties function very similarly across the two versions. For this reason, we combined the samples for the two versions in the remaining scale analyses. Similarly, no meaningful DIF was found by setting, gender, and age. The R^2 values were well below 0.02 for combined DIF types (uniform and nonuniform). These results support the feasibility of assessment of these items in PAC settings regardless of setting, gender, and age.

Known Groups Validity

Comparing the performance of patients/residents on the PROMIS Depression data elements to other patient/resident characteristics adds information about the validity of the data elements. If known or logical associations between patients/resident characteristics and data elements are

observed in data from the National Beta Test, this contributes to the evidence that the data elements are valid, or assessing the construct that they are intended to capture.

Table 4.3 shows the PROMIS Depression T-score for the overall admission sample stratified by patient/resident characteristics and clinical groups as described in Chapter 1: gender (male or female as documented by National Beta Test assessor), age (18–44, 45–64, 65–74, 75–89, or 90 and over), length of stay (in days), disposition at discharge (e.g., to another PAC setting, home, to hospital), sepsis, heart failure, stroke, and two ADLs: toileting (not available for HHA patients) and ability to transfer from lying to sitting. As a reminder, these clinical conditions were chosen based on their common occurrence across settings, their frequent relationship with many of the data elements tested in the National Beta Test, and their availability in all four settings (i.e., equivalent information was collected on the OASIS, IRF-PAI, LCDS, and MDS). Setting-specific results for these individual characteristics are presented in the appendix in Tables A.15–A.18.

Table 4.3. Mean (SD) PROMIS Depression T-Score by Patient/Resident Characteristics and Clinical Groups

Patient/Resident Characteristics and Clinical Groups	PROMIS Depression T-Score M (SD)
Gender (*n* = 2,884[a])	
Male (*n* = 1,196)	50.5 (9.7)
Female (*n* = 1,688)	51.6 (9.7)
Age (*n* = 2,873[a])	
18–44 (*n* = 39)	54.0 (11.3)
45–64 (*n* = 304)	54.6 (10.4)
65–74 (*n* = 902)	51.6 (9.7)
75–89 (*n* = 1,322)	50.2 (9.4)
90+ (*n* = 306)	49.8 (9.2)
Length of stay (*n* = 2,632)	Pearson *r* = 0.005
Disposition at discharge (*n* = 2,834[a])	
Home (*n* = 1,328)	50.2 (9.3)
Hospital (*n* = 196)	52.0 (10.5)
Hospice (*n* = 40)	50.1 (8.5)
HHA (*n* = 619)	51.7 (9.6)
IRF (*n* = 50)	54.2 (11.2)
LTCH (*n* = 12)	48.6 (10.2)
SNF (*n* = 277)	52.3 (10.4)
Other (*n* = 312)	51.9 (9.9)
Clinical conditions (*n* = 2,220)	
Sepsis	

Patient/Resident Characteristics and Clinical Groups	PROMIS Depression T-Score M (SD)
Yes (*n* = 148[a])	53.5 (10.8)
No (*n* = 2,072)	51.1 (9.7)
Heart failure	
Yes (*n* = 379)	50.9 (9.8)
No (*n* = 1,841)	51.3 (9.7)
Stroke	
Yes (*n* = 192)	51.2 (10.4)
No (*n* = 2,028)	51.2 (9.7)
Hygiene—Toileting (*n* = 1,496[a])	
Independent (*n* = 336)	52.9 (10.3)
Setup or clean-up assistance (*n* = 77)	48.2 (8.2)
Supervision or touching assistance (*n* = 319)	49.6 (9.0)
Partial/moderate assistance (*n* = 360)	51.6 (9.8)
Substantial/maximal assistance (*n* = 332)	52.7 (9.8)
Dependent (*n* = 336)	52.9 (10.3)
Mobility—Lying to sitting (*n* = 1,856[a])	
Independent (*n* = 192)	50.2 (9.8)
Setup or clean-up assistance (*n* = 112)	50.5 (9.1)
Supervision or touching assistance (*n* = 527)	49.8 (9.1)
Partial/moderate assistance (*n* = 606)	51.1 (9.6)
Substantial/maximal assistance (*n* = 289)	52.5 (10.2)
Dependent (*n* = 130)	53.7 (10.0)

[a] Indicates significant ($p < 0.05$) association between patient/resident characteristic and PROMIS Depression T-Score.

Although we did not have hypotheses or expectations for all characteristics and conditions listed in this table, we did expect depression to be related to gender, age, stroke, toileting, and mobility (ability to transfer from lying to sitting), such that patients/residents with higher PROMIS Depression T-score would tend to be female,[72] younger,[73] to have had a stroke,[74] and to have less independence in ADLs.[75]

In the overall sample, the PROMIS Depression T-score is significantly associated with gender, age, disposition at discharge, sepsis, and the two ADLs. Not surprisingly, these associations are generally consistent with the pattern of associations found with positive screens on the PHQ-2.

[72] Djernes, 2006; Jones, Marcantonio, and Rabinowitz, 2003.

[73] Jones, Marcantonio, and Rabinowitz, 2003.

[74] Jongenelis et al., 2004.

[75] Kaup et al., 2007.

Gender and Age

- Gender, overall, is significantly associated with the PROMIS Depression T-score, ($t_{(2882)}$ = 2.9, $p < 0.05$), such that women scored higher (M = 51.6, SD = 9.7) than men (M = 50.5, SD = 9.7). A similar trend was observed at the setting level in HHAs but not in IRFs, LTCHs, or SNFs. Results are similar to PHQ-2 to 9 findings, where women were more likely to screen positive on the PHQ-2. *We expected this association, which reflects the incidence of depression in the broader population. This finding in the overall sample supports the validity of the PROMIS depression data elements to accurately assess for depression.*
- Age, overall, is significantly associated with PROMIS Depression T-score ($F_{(4,2868)}$ = 15.51, $p < 0.01$), such that the younger age groups (i.e., 18–44, 45–64, and 65–74) tend to have significantly higher scores, or more depression (M [SD] = 54.0 [11.3], 54.6 [10.4], and 51.6 [9.7], respectively) than the 90+ age group (M [SD] = 49.8 [9.2]; $t_{(2868)}$ = 2.52, 6.14, and 2.71, respectively, $p < 0.05$), except for the 75–89 age group (M [SD] = 50.2 [9.4]), which is not significantly different from the 90+ group. A similar trend is observed at the setting level, where younger age groups in general have higher PROMIS Depression T-scores, indicating more depression. The 45–64 age group has the highest score in all settings except for SNFs, where those between 18–44 are the highest-scoring group (M = 57.4, SD = 13.7). Overall results are consistent with PHQ-2 to 9 findings, where the younger groups (e.g., 45–64) have higher PHQ-9 summary scores. *We expected this association, which reflects the incidence of depression in the broader population. These findings support the validity of the PROMIS depression data elements to accurately assess for depression.*

Length of Stay, Disposition at Discharge

- Length of stay, overall (M = 21.6 days, SD = 12.8), is not significantly associated with PROMIS Depression T-score (Pearson correlation $r = 0.005$, $p > 0.05$). At the setting level, length of stay is also not associated with the T-scores. These results mirror PHQ-2 to 9 findings. *However, we did not expect to observe an association with length of stay.*
- Overall, disposition at discharge is significantly associated with PROMIS Depression T-score ($F_{(7,2826)}$ = 4.24, $p < 0.001$). Relative to patients/residents discharged to all other placements, those discharged from any of the four settings to LTCHs have the lowest T-score (M = 48.6, SD = 10.2), indicating less severe depressive symptoms, and those discharged from any of the four settings to IRFs have the highest T-score (M = 54.2, SD = 11.2), indicating higher depressive symptom severity. There are no significant associations at the setting level, however. These results, overall, are similar to PHQ-2 to 9 findings, such that, compared with discharges to all other placements, patients/residents discharged to IRFs were more likely to screen positive on PHQ-2, whereas patients/residents discharged to LTCHs were less likely to screen positive on PHQ-2. *However, we did not expect to observe an association with discharge disposition.*

Clinical Conditions

- There is an overall significant difference in PROMIS Depression T-scores ($t_{(2218)}$ = 2.83, $p < 0.05$) between patients with sepsis (M = 53.5, SD = 10.8) and those without (M = 51.1, SD = 9.7), similar to the PHQ-2 to 9 findings, where those with sepsis were more

likely to screen positive on the PHQ-2. A similar trend is observed in LTCHs, where patients with sepsis have a higher T-score (M = 56.5, SD = 11.3) than those without (M = 53.3, SD = 9.8; $t_{(385)} = 2.42$, $p < 0.05$). In HHAs, however, the mean difference is reversed, such that those with sepsis have a significantly lower mean (M = 43.2, SD = 6.8) than those without (M = 49.6, SD = 9.0; $t_{(414)} = 1.98$, $p < 0.05$). There is no such mean difference observed in IRFs or SNFs. *We did not have any expectations about the relationship between PROMIS Depression T-scores and sepsis.*

- There is no association of heart failure with PROMIS Depression T-scores overall or in any of the settings. *We did not expect an association with this variable.*
- There is no association of stroke with PROMIS Depression T-scores overall or in any of the settings. *We predicted an association between the PROMIS Depression T-scores and stroke. However, finding no association is consistent with the lack of a finding for PHQ-2 screening.*

ADLs: Toileting and Ability to Transfer from Lying to Sitting

- Similar to PHQ-2 to 9 findings, where those with greater dependence were more likely to screen positive, overall level of assistance with toileting hygiene is significantly associated with PROMIS Depression T-score ($F_{(5,1490)} = 6.78$, $p < 0.001$), such that scores tend to increase with increased need for assistance. For example, those needing minimal assistance (i.e., setup or clean-up assistance) have lower scores (M = 48.2, SD = 8.2) than those requiring substantial/maximal assistance (M = 52.7, SD = 9.8) and those who are completely dependent (M = 52.9, SD = 10.3). However, this trend does not apply to patients/residents characterized as completely independent for toileting hygiene. This group has average depression T-scores similar to those who are completely dependent (M = 52.9, SD = 10.3). By setting, toileting hygiene is significantly associated with PROMIS Depression T-score only in IRFs, such that those who need more assistant in general have higher PROMIS Depression T-scores. *These findings conform to our expectations that patients/residents with less ADL independence would have higher PROMIS Depression T-scores, which we found to be generally true in the overall sample and in IRFs but not in other settings.*
- Overall ability to transfer from lying to sitting is significantly associated with PROMIS Depression T-score ($F_{(5,1850)} = 5.57$, $p < 0.001$), such that those who need supervision or touching assistance have the lowest T-score (M = 49.8, SD = 9.1) as compared with those who are completely dependent (M = 53.7, SD = 10). These results are consistent with PHQ-2 to 9, where those who were completely dependent were more likely to screen positive. By setting, mobility is only significantly associated with PROMIS Depression T-score in IRFs and LTCHs, such that those who need more assistance in general have higher PROMIS Depression T-scores. *As with toileting, these findings conform to our expectations that patients/residents with less ADL independence would have higher PROMIS Depression T-scores, which we found to be generally true in the overall sample and in IRFs and LTCHs but not in other settings.*

Time to Complete

Table 4.4 shows, on average, time to complete the PROMIS Depression data elements overall and by setting for each version separately and combined across versions. On average, the

time to complete for version A (past 7 days) was 2.2 minutes (SD = 0.8) and ranged from 2.0 minutes (SD = 0.9) in HHAs to 2.3 minutes (SD = 0.8) in LTCHs. For version B (past 3 days), time to complete on average was also 2.2 minutes (SD = 0.8) and ranged from 1.9 minutes (SD = 0.8) in IRFs to 2.4 minutes in both HHAs (SD = 0.8) and SNFs (SD = 0.9). Time to complete was significantly associated with setting type using version A ($F_{(3,737)} = 2.73$, $p < 0.05$), but none of the pair-wise comparisons among the settings is statistically significant ($p > 0.05$). Time to complete was also significantly associated with setting type using version B ($F_{(3,782)} = 11.1$, $p < 0.01$), such that it took significantly less time to complete in IRFs than in HHAs ($t_{(782)} = 4.99$, $p < 0.01$) and SNFs ($t_{(782)} = 5.03$, $p < 0.01$).

Table 4.4. Time to Complete for PROMIS Depression Data Elements (minutes)

		HHA (*n* = 368)	IRF (*n* = 478)	LTCH (*n* = 250)	SNF (*n* = 431)	Overall (*n* = 1,527)
Past 7 days	Mean (SD)	2.0 (0.9)	2.2 (0.8)	2.3 (0.8)	2.2 (0.9)	2.2 (0.8)
Past 3 days	Mean (SD)	2.4 (0.8)	1.9 (0.8)	2.2 (0.7)	2.4 (0.9)	2.2 (0.8)
Combined	Mean (SD)	2.2 (0.9)	2.1 (0.8)	2.2 (0.7)	2.3 (0.9)	2.2 (0.8)

Time to complete was also evaluated according to urbanicity (urban versus nonurban), geographic region (Northeast, South, Midwest, West), facility ownership (for-profit versus nonprofit), and facility size (above or below setting-type median) to evaluate the generalizability of these performance results (see Tables A.19–A.22 in the appendix). No significant differences were found for time to complete the PROMIS Depression items in these sensitivity analyses.

Interrater Reliability

Table 4.5 shows kappa interrater reliability coefficients for the two versions combined overall and by setting. As described in more detail in Volume 3, paired assessment data for interrater reliability evaluation were collected on a subset of the National Beta Test communicative admission sample of patients/residents according to setting-level target totals. For example, each participating LTCH was asked to conduct 20 paired assessments to contribute to interrater reliability. Inclusion in interrater reliability data collection depended on paired facility/staff and research nurse assessors' ability to schedule assessments. Kappas were computed on 935 patients/residents who were assessed by research nurse and facility/agency staff assessor pairs: 192 in HHAs, 250 in IRFs, 229 in LTCHs, and 264 in SNFs. Kappas for the PROMIS Depression data elements were excellent, ranging from 0.97 to 0.99. At the setting level, kappas were also excellent, ranging from 0.98 to 1.00 in HHAs, 0.97 to 0.99 in IRFs, 0.97 to 1.00 in LTCHs, and 0.96 to 0.99 in SNFs. Generally speaking, kappas are all excellent and there is little variation among them at the setting level.

Table 4.5. Interrater Reliability Kappa or Weighted Kappa for PROMIS Depression Data Elements by Setting

Data Element	HHA (*n* = 192)	IRF (*n* = 250)	LTCH (*n* = 229)	SNF (*n* = 264)	Overall (*n* = 935)
How often felt worthless (e2a)	0.99	0.99	0.99	0.98	0.99
How often felt there was nothing to look forward to (e2b)	0.99	0.98	0.97	0.96	0.97
How often felt helpless (e2c)	0.98	0.98	1.00	0.98	0.98
How often felt sad (e2d)	0.98	0.97	0.99	0.99	0.98
How often felt lonely (e2e)	1.00	0.98	0.98	0.99	0.99
How often felt depressed (e2f)	1.00	0.99	0.98	0.97	0.98
How often felt there was no reason for living (e2g)	1.00	0.97	0.97	0.97	0.98
How often felt hopeless (e2h)	0.99	0.99	0.98	0.97	0.98
Depression total[a]	1.00	0.99	1.00	0.99	0.99

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

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.23–A.26 in the appendix). No significant differences were found for interrater reliability of the PROMIS Depression data elements in these sensitivity analyses.

Table 4.6 shows the percent agreement for the PROMIS Depression data elements overall and by setting. Overall, percent agreement was high for all data elements, ranging from 98 to 99 percent with minimal setting differences.

Table 4.6. Interrater Reliability—Percent Agreement for PROMIS Depression Data Elements

Data Element	HHA (*n* = 192)	IRF (*n* = 250)	LTCH (*n* = 229)	SNF (*n* = 264)	Overall (*n* = 935)
How often felt worthless (e2a)	99	98	99	98	99
How often felt there was nothing to look forward to (e2b)	99	98	98	98	98
How often felt helpless (e2c)	99	98	100	98	99
How often felt sad (e2d)	99	98	99	98	99
How often felt lonely (e2e)	100	99	99	98	99
How often felt depressed (e2f)	100	99	99	97	99
How often felt there was no reason for living (e2g)	100	99	99	98	99
How often felt hopeless (e2h)	99	99	99	98	99

Scale Reliability

Because the PROMIS Depression data elements are typically interpreted at the scale level, we conducted additional analyses to evaluate the unidimensionality and internal-consistency reliability of these data elements. These analyses provided strong support for the scalability of the data elements. Item total correlations ranged from 0.70 to 0.83, with an average data element–item correlation of 0.58 and a Cronbach's alpha of 0.92. The explained common variance, which has a maximum of 1, is 0.83 for this set of data elements, indicating strong unidimensionality.

Admission to Discharge

Table 4.7 summarizes patterns of change from admission to discharge on the PROMIS Depression data elements and total and T-scores for data combined from both versions. As described in more detail in Volume 3, discharge data were collected on a subset of the National Beta Test admission sample of patients/residents. Availability of discharge data depended on advance notification of discharge and the ability to schedule assessments among the facility staff assessors in each participating site. Patterns are characterized as "no change" (scores stay the same at admission and discharge), "improve" (scores reflect improvement from admission to discharge), and "worsen" (scores indicate a decline from admission to discharge). Results are based on data from 779 patients/residents, 461 of whom received version A and 318 of whom received version B. Overall, responses demonstrated some degree of stability from admission to discharge with an overall range of 50 percent (felt sad) to 84 percent (felt no reason for living) experiencing no change, depending on the particular data element. However, compared with admission, patients/residents at discharge were more likely to show improvement in symptoms than to worsen. For example, patients/residents were less likely to feel worthless at discharge than at admission ($t_{(773)} = 8.29$, $p < 0.001$). Across all the data elements, the percentage worsening ranged from 5 to 16 percent, and the percentage showing improvement ranged from 11 to 55 percent. Further, similar to results for the PHQ, over half of all patients/residents showed improvement from admission to discharge in PROMIS Depression total score and T-score. There was little noticeable variation across settings.

Table 4.7. Admission to Discharge Results for PROMIS Depression Data Elements (percent)

Data Element	HHA (*n* = 145)	IRF (*n* = 331)	LTCH (*n* = 81)	SNF (*n* = 222)	Overall (*n* = 779)
How often felt worthless (e2a)					
No change	66	66	65	74	68
Worsen	8	6	13	9	8
Improve	26	28	23	17	24
How often felt there was nothing to look forward to (e2b)					
No change	74	65	64	71	68
Worsen	5	9	17	8	8
Improve	21	27	19	21	23
How often felt helpless (e2c)					
No change	60	51	44	55	53
Worsen	11	12	22	10	12
Improve	29	37	33	35	35
How often felt sad (e2d)					
No change	57	45	45	54	50
Worsen	11	16	16	11	14
Improve	32	38	39	35	36
How often felt lonely (e2e)					
No change	63	58	50	55	57
Worsen	14	14	24	18	16
Improve	23	29	26	27	27
How often felt depressed (e2f)					
No change	70	58	54	57	60
Worsen	6	11	14	14	11
Improve	24	31	32	29	30
How often felt there was no reason for living (e2g)					
No change	83	85	72	86	84
Worsen	4	4	10	6	5
Improve	13	11	19	8	11
How often felt hopeless (e2h)					
No change	80	75	59	74	74
Worsen	5	7	15	9	8
Improve	15	19	26	16	18
Depression total score					
No change	28	23	19	24	24
Worsen	18	21	31	24	22

Data Element	HHA (*n* = 145)	IRF (*n* = 331)	LTCH (*n* = 81)	SNF (*n* = 222)	Overall (*n* = 779)
Improve	54	57	51	52	54
Depression T-score					
No change	22	18	13	23	20
Worsen	21	24	35	26	25
Improve	57	58	52	51	55

The summed score to T-score conversion table for the eight PROMIS Depression data elements administered in the National Beta Test is included as Table A.27 in the appendix.

Assessor Feedback

Facility staff who participated in the assessor survey considered PROMIS Depression to be only somewhat clinically useful relative to all other data elements. Research nurses who participated in the assessor survey gave PROMIS Depression a higher rating for clinical utility than did facility staff, with an average rating falling within the middle range of scores received for all data elements. Survey participants also deemed PROMIS Depression to be more burdensome to collect than other data elements, with facility staff ratings falling in the high range and research nurse ratings falling in the middle to high range for clinical burden, relative to all other data elements. Respondents in both groups also rated PROMIS Depression higher than other data elements for burden to the patient/resident. When survey participants were asked about factors impacting data collection, they noted that patient confusion and discomfort in answering these questions caused them to question the accuracy of the responses. These factors likely also contributed to the higher assessment of burden for this data element.

Facility staff in the focus groups noted some challenges to administering PROMIS Depression, echoing findings from the assessor survey. Facility staff in focus groups observed that some patients/residents got uncomfortable upon hearing the introductory language, which they believed assumes that the interviewee was experiencing emotional distress ("I'm now going to ask you about your emotional distress"). As such, some facility staff admitted omitting the word "your" to avoid these negative reactions. Focus group participants also thought the questions could be upsetting to the patient/resident, forcing some to confront depression symptoms (e.g., "nothing to look forward to" and "no reason for living"). Facility staff noted the additional time and resources that would be necessary to provide appropriate follow-up care and empathy when asking the PROMIS Depression data elements, which likely contributed to perceptions of assessor burden as reflected in survey findings. Facility staff in focus groups reported being uncomfortable asking about the topics covered by PROMIS Depression data elements during data collection without taking time to provide appropriate follow-up care and empathy.

Regarding the questions themselves, facility staff reported the need to differentiate "helpless" from "hopeless." Patient/residents often thought of these and other words used in the prompts as synonyms, which also caused confusion and frustration:

> Helpless, sad, lonely, depressed, . . . all those words are kind of the same meaning. . . . And the more words and more things you throw at them just makes it more complex.
>
> —Durham, SNF staff

> Sometimes people don't get . . . the difference between hopeless and helpless. So for them it's just like why do you keep asking me this same question over and over again? I've already told you. I don't feel down or depressed or sad or hopeless or helpless. That's why it's so repetitive to them is because they view those words as synonyms even though they're not.
>
> —Nashville, SNF staff

Facility staff in focus groups were also uncertain about the validity of patient/resident responses to the PROMIS Depression data elements. Some staff noted that patients/residents seemed to be answering "no" to all the questions without considering the content, possibly because they wanted to get through the list of data elements quickly because of fatigue or discomfort. Another potential but not unexpected threat to validity was the presence of a spouse or other family member during the assessment: Answers to the PROMIS depression data elements appeared to be more positive and/or the spouse might directly contradict the patient's/resident's answer. Facility staff's perceptions that patients/residents were not completely forthcoming in their answers moderated their enthusiasm for considering PROMIS Depression as part of a standardized patient assessment.

Participants in the focus groups did note some strengths. First, facility staff in the focus groups noted that assessment of depression was clinically relevant to them. A specific strength of the PROMIS Depression data element was that the alternate wordings of several questions can help identify when a patient/resident needs additional assessment without the patient/resident needing to self-identify as depressed:

> Some of the anxiety/depression questions were helpful because they were a new way of looking at being depressed. You might not call it depressed but the fact like if they couldn't sleep or some of these, they felt like they didn't want to do anything or they had nothing to look forward to, they might not interpret that exactly as depression.
>
> —Boston, HHA staff

In summary, similar to the PHQ, standardizing depression assessments was important to the assessors in both the survey and focus group discussions. In focus groups, the research nurses did not offer feedback specific to the PROMIS Depression data elements, and facility staff had mixed feedback on this set of data elements.

Summary

Results for PROMIS Depression data elements indicate moderate overall support for cross-setting standardization. Assessors considered PROMIS Depression data elements to be somewhat clinically useful but to have relatively high data collection burden and burden to the patient/resident. Assessors noted that some questions were upsetting to the patient/resident, suggesting that these data elements will require staff training to improve comfort with administering these data elements and responding to patient/resident need afterward. Assessors were also concerned that patients/residents were hesitant to disclose information about their mood and thus perhaps were not completely forthcoming in their answers. As with the PHQ-2 to 9, for the most part associations of PROMIS Depression T-scores with patient/resident characteristics and clinical conditions provided support for the validity of the assessment. We also found that the data elements performed similarly according to setting, gender, and age (i.e., did not observe any DIF), which supports their use in the PAC population. For interrater reliability, kappas were excellent and percent agreement was high, with little variation across the two versions (past 7 days versus past 3 days) or at the setting level. An important finding is that responses demonstrated some degree of stability from admission to discharge, with patients/residents more likely to show improvement in symptoms than to worsen at discharge. The combined results for PROMIS Depression data elements are somewhat mixed, showing excellent interrater reliability but only moderate feasibility, especially because of the relatively high burden, and moderate clinical utility as a candidate data element for standardization across PAC settings.

5. PROMIS Anxiety

Data Element Description

The PROMIS Anxiety data elements include eight data elements from the PROMIS Anxiety Item Bank, which has a total of 29 items that assess self-reported panic, fear, worry, ability to calm down or focus, and nervousness. These data elements are completed through patient/resident interview and are not currently included in any PAC patient/resident assessment instruments. In the National Beta Test, the PROMIS Anxiety data elements were collected in two versions. One version was identical to the format in the PROMIS Anxiety Item Bank using a reference to "the past 7 days" for symptoms. The second version, shown in Figure 5.1, was identical except that it referenced "the past 3 days" to align more closely with PAC assessment data elements that are currently in use.

Figure 5.1. PROMIS Anxiety Data Elements

SAY TO PATIENT/RESIDENT: "I am now going to ask you about your emotional distress, specifically anxiety and how you have been feeling over the past 3 days." I will also ask about some common problems that sometimes go along with feeling anxious. This is not meant to give you a diagnosis. Some of the questions might seem personal, but all patients/residents are asked to answer them. Knowing the answers to these questions will help us provide you with a more individualized care plan."
E3a. ASK PATIENT/RESIDENT: "In the past 3 days, I felt worried:" ☐ 1 = Never ☐ 2 = Rarely ☐ 3 = Sometimes ☐ 4 = Often ☐ 5 = Always ☐ 7 = **Patient/resident declined to respond** ☐ 9 = **Unknown or unable to assess**
E3b. ASK PATIENT/RESIDENT: "In the past 3 days, my worries overwhelmed me:**"** ☐ 1 = Never ☐ 2 = Rarely ☐ 3 = Sometimes

- ☐ 4 = Often
- ☐ 5 = Always
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

E3c. ASK PATIENT/RESIDENT:
"In the past 3 days, I had trouble paying attention:"

- ☐ 1 = Never
- ☐ 2 = Rarely
- ☐ 3 = Sometimes
- ☐ 4 = Often
- ☐ 5 = Always
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

E3d. ASK PATIENT/RESIDENT:
"In the past 3 days, I felt nervous:"

- ☐ 1 = Never
- ☐ 2 = Rarely
- ☐ 3 = Sometimes
- ☐ 4 = Often
- ☐ 5 = Always
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

E3e. ASK PATIENT/RESIDENT:
"In the past 3 days, I had difficulty calming down:"

- ☐ 1 = Never
- ☐ 2 = Rarely
- ☐ 3 = Sometimes
- ☐ 4 = Often
- ☐ 5 = Always
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unknown or unable to assess**

E3f. ASK PATIENT/RESIDENT:
"In the past 3 days, I found it hard to focus on anything other than my anxiety:"

- ☐ 1 = Never
- ☐ 2 = Rarely

☐ 3 = Sometimes ☐ 4 = Often ☐ 5 = Always ☐ 7 = **Patient/resident declined to respond** ☐ 9 = **Unknown or unable to assess**
E3g. ASK PATIENT/RESIDENT: "In the past 3 days, I felt like I needed help for my anxiety:" ☐ 1 = Never ☐ 2 = Rarely ☐ 3 = Sometimes ☐ 4 = Often ☐ 5 = Always ☐ 7 = **Patient/resident declined to respond** ☐ 9 = **Unknown or unable to assess**
E3h. ASK PATIENT/RESIDENT: "In the past 3 days, I had sudden feelings of panic:" ☐ 1 = Never ☐ 2 = Rarely ☐ 3 = Sometimes ☐ 4 = Often ☐ 5 = Always ☐ 7 = **Patient/resident declined to respond** ☐ 9 = **Unknown or unable to assess**

Testing Objectives

As stated above, two versions of the PROMIS Anxiety data elements were administered during the National Beta Test: a standard (past 7 days) version conforming to the format used in the PROMIS Anxiety item library, and a modified (past 3 days) version. The 7-day version was administered in Market Group A and the 3-day version in Market Group B. For more details on Market Group A and B samples and characteristics, please refer to Volume 3. Basic descriptive statistics (e.g., frequencies and scale scores) are first presented for admission data to characterize the responses to the PROMIS Anxiety data elements according to version for patients/residents in each setting and for the overall sample. Next, we present differences in the versions that we evaluated by (1) conducting t-tests to compare data element and scale level means across the two versions (using Cohen's *d* value of greater than 0.2 to indicate meaningful version differences)

and (2) conducting DIF by version to determine whether data elements functioned differently within the scale according to version.

Because we saw little evidence of performance differences according to version, the majority of results, with the exception of feasibility results (data element frequencies, rates of missingness, and time to complete), are presented for the two versions combined. Specifically, we used the combined admission response data to (1) evaluate DIF according to setting, gender, and age, (2) examine PROMIS Anxiety T-scores by patient/resident characteristics and clinical groups of interest (i.e., known groups validity), and (3) evaluate interrater reliability (kappa and percent agreement), data element-total correlations, Cronbach's alpha internal consistency values, and unidimensionality indices. Also, frequencies at admission and discharge were compared for the combined versions to inform stability or possible change over time. Finally, because these data elements come from the IRT-calibrated PROMIS Anxiety Item Bank, we provide a summed score to T-score conversion table as Table A.42 the appendix. This approach allows for the calculation of T-scores with M = 50 and SD = 10, allowing results to be directly compared with general population PROMIS Anxiety values.

Results

Feasibility

Frequencies/Missing

Tables 5.1 and 5.2 show the percentage of responses for each PROMIS Anxiety data element, both overall and by setting. Table 5.1 presents frequencies for Market Group A ("past 7 days"), and Table 5.2 presents frequencies for Market Group B ("past 3 days"). Version A was administered to 1,492 patients/residents: 228 in HHAs, 483 in IRFs, 249 in LTCHs, and 532 in SNFs. Version B was administered to 1,479 patients/residents: 409 in HHAs, 287 in IRFs, 220 in LTCHs, and 563 in SNFs. Overall, more than 96 percent of the sample was administered one of the two versions. Among these patients/residents, overall missing data at the data element level ranged from 0.60 percent to 1.81 percent in version A and 0.41 percent to 1.56 percent in version B with minimal setting differences.

The tables present percentages of patients/residents endorsing frequencies of symptoms from "never" to "always." Across the two market groups, there are minimal differences, and in general, patients/residents tended to most frequently endorse the "never" response category. Across the settings, patients/residents from LTCHs using version B seemed to endorse the highest frequency category (i.e., always) consistently more than those using version A and those from the other settings. For example, among those from LTCHs, 6 percent of respondents using version A and 10 percent using version B endorsed the "always" category in response to the "felt worried" item. In comparison, endorsement of this category in versions A and B respectively is 4 percent and 2 percent from HHAs, 5 percent and 3 percent from IRFs, and 6 and 3 percent from

SNFs. The tables also show the mean of the total Anxiety score (range from 0 to 40) and the T-score (a standardized metric normed to the general population with M = 50 and SD = 10), overall and by setting across two different versions. The same pattern is observed for the total and the T-scores as in individual data elements. That is, there is little difference between the two versions, but similar to rates for the depression data elements, LTCH patients consistently have higher anxiety scores (total scores of 16.3 and 17.0 for versions A and B, respectively; T-scores of 53.4 and 53.8 for versions A and B, respectively) than patients/residents from the other settings (total scores averaging approximately 14 and T-scores averaging approximately 50).

Table 5.1. Overall and Setting-Specific Response Frequencies for PROMIS Anxiety Data Elements at Admission: Market Group A (percent)

Data Element	HHA (*n* = 228)	IRF (*n* = 483)	LTCH (*n* = 249)	SNF (*n* = 532)	Overall (*n* = 1,492)
How often: felt worried (e3a)					
Never	35	27	30	41	34
Rarely	22	21	15	14	18
Sometimes	29	32	35	31	32
Often	10	15	13	9	12
Always	4	5	6	6	5
How often: overwhelmed by worries (e3b)					
Never	59	61	52	68	62
Rarely	18	14	14	13	14
Sometimes	14	15	22	14	16
Often	5	6	8	3	5
Always	3	3	4	2	3
How often: trouble paying attention (e3c)					
Never	57	46	48	58	52
Rarely	20	21	16	12	17
Sometimes	17	23	28	23	23
Often	4	9	6	5	6
Always	1	2	3	2	2
How often: felt nervous (e3d)					
Never	45	40	35	51	44
Rarely	20	22	19	16	19
Sometimes	24	26	35	25	27
Often	9	9	5	4	7
Always	2	3	5	4	4

How often: difficulty calming down (e3e)					
Never	65	59	50	67	61
Rarely	19	18	19	15	17
Sometimes	13	17	19	12	15
Often	3	4	8	4	5
How often: focused on anxiety (e3f)					
Never	64	64	55	69	64
Rarely	22	14	18	13	16
Sometimes	11	13	19	11	13
Often	2	6	6	5	5
Always	1	3	2	1	2
How often: felt need for help with anxiety (e3g)					
Never	67	64	50	72	65
Rarely	15	14	18	9	13
Sometimes	15	12	17	12	13
Often	2	8	10	4	6
Always	2	4	5	3	3
How often: sudden feelings of panic (e3h)					
Never	80	70	61	78	73
Rarely	7	15	14	9	12
Sometimes	10	9	16	9	11
Often	2	4	6	2	3
Always	0	1	2	1	1
Anxiety total Mean (SD)	13.8 (6.1)	15.1 (6.9)	16.3 (7.2)	13.7 (6.3)	14.6 (6.7)
Anxiety T-score Mean (SD)	49.8 (9.5)	51.7 (9.9)	53.4 (10.3)	49.2 (9.8)	50.8 (10.0)

Table 5.2. Overall and Setting-Specific Response Frequencies for PROMIS Anxiety Data Elements at Admission: Market Group B (percent)

Data Element	HHA (*n* = 409)	IRF (*n* = 287)	LTCH (*n* = 220)	SNF (*n* = 563)	Overall (*n* = 1,479)
How often: felt worried (e3a)					
Never	41	41	25	38	37
Rarely	22	15	17	21	19
Sometimes	27	30	33	29	29
Often	8	10	15	9	10
Always	2	3	10	3	4
How often: overwhelmed by worries (e3b)					
Never	68	68	53	69	66
Rarely	12	15	14	11	13
Sometimes	13	10	18	12	13
Often	6	5	12	5	6
Always	1	2	4	2	2
How often: trouble paying attention (e3c)					
Never	55	51	46	55	53
Rarely	17	15	20	15	17
Sometimes	20	26	20	23	22
Often	6	5	8	5	6
Always	1	2	6	2	2
How often: felt nervous (e3d)					
Never	49	45	40	48	46
Rarely	21	15	17	18	18
Sometimes	21	29	25	24	24
Often	8	7	13	8	8
Always	2	5	5	3	3
How often: difficulty calming down (e3e)					
Never	62	61	50	67	62
Rarely	21	16	15	13	16
Sometimes	12	18	22	14	15
Often	4	4	9	5	5
How often: focused on anxiety (e3f)					
Never	68	65	54	68	65
Rarely	13	14	18	14	14
Sometimes	13	15	15	12	13
Often	4	3	8	5	5

Data Element	HHA (*n* = 409)	IRF (*n* = 287)	LTCH (*n* = 220)	SNF (*n* = 563)	Overall (*n* = 1,479)
Always	1	3	6	1	2
How often: felt need for help with anxiety (e3g)					
Never	71	64	49	68	65
Rarely	9	13	13	11	11
Sometimes	14	15	22	12	15
Often	5	4	10	6	6
Always	2	3	6	3	3
How often: sudden feelings of panic (e3h)					
Never	77	76	59	80	75
Rarely	13	10	15	8	11
Sometimes	8	10	13	8	9
Often	2	3	9	4	4
Always	1	1	4	1	1
Anxiety total Mean (SD)	13.6 (6.1)	14.3 (6.4)	17.0 (8.1)	13.9 (6.5)	14.4 (6.7)
Anxiety T-score Mean (SD)	49.1 (9.7)	50.0 (10.0)	53.8 (11.5)	49.6 (9.8)	50.2 (10.2)

Version Comparison and Differential Item Functioning

Results of the comparison of data element and scale means by version are shown in Table A.28 in the appendix. While two data elements ("I felt worried," "My worries overwhelmed me") out of eight showed statistically significant differences by version, these were below the Cohen's *d* value of 0.2. At the scale level, these differences were not significant and Cohen's *d* values were well below 0.2. In addition, results from DIF evaluation according to version showed R^2 effect sizes well below the 0.02 cut-off for nontrivial DIF for all data elements (Table A.29 in the appendix). This means that the data element properties function very similarly in IRT models regardless of version. For this reason, we combined the samples in the remaining scale analyses. Similarly, no meaningful DIF was found by setting, gender, or age. All R^2 values were well below 0.02. These results support the feasibility of assessment of these data elements in PAC settings regardless of setting, gender, and age.

Known Groups Validity

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

Table 5.3 shows the PROMIS Anxiety T-score for the overall admission sample (combined across versions) 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 (18–44, 45–64, 65–74, 75–89, or 90 and over), length of stay (in days), disposition at discharge (e.g., to another PAC setting, home, to hospital), sepsis, heart failure, stroke, and two ADLs: toileting (not available for HHA patients) and ability to transfer from lying to sitting. As a reminder, these clinical conditions were chosen based on their common occurrence across settings, their frequent relationship with many of the data elements tested in the National Beta Test, and their availability in all four settings (i.e., equivalent information was collected on the OASIS, IRF-PAI, LCDS, and MDS). Setting-specific results for these individual characteristics are presented in the appendix in Tables A.30–A.33.

Table 5.3. Mean (SD) PROMIS Anxiety T-Score by Patient/Resident Characteristics and Clinical Groups

Patient/Resident Characteristics and Clinical Groups	PROMIS Anxiety T-Score M (SD)
Gender (*n* = 2,870[a])	
Male (*n* = 1,188)	49.7 (10.1)
Female (*n* = 1,682)	51.0 (10.0)
Age (*n* = 2,859[a])	
18–44 (*n* = 39)	52.0 (11.3)
45–64 (*n* = 303)	54.6 (12.3)
65–74 (*n* = 900)	51.5 (10.2)
75–89 (*n* = 1,315)	49.3 (9.6)
90+ (*n* = 302)	48.5 (9.1)
Length of stay (*n* = 2,632)	Pearson *r* = –0.02
Disposition at discharge (*n* = 2,821[a])	
Home (*n* = 1,324)	49.6 (9.7)
Hospital (*n* = 194)	51.2 (10.6)
Hospice (*n* = 40)	49.7 (10.4)
HHA (*n* = 618)	50.7 (10.0)
IRF (*n* = 51)	53.3 (11.5)
LTCH (*n* = 11)	49.3 (10.5)
SNF (*n* = 273)	52.3 (10.5)
Other (*n* = 310)	51.4 (10.2)26
Clinical conditions (*n* = 2,209)	
Sepsis	
Yes (*n* = 147[a])	53.8 (11.2)
No (*n* = 2,062)	50.4 (9.9)

Patient/Resident Characteristics and Clinical Groups	PROMIS Anxiety T-Score M (SD)
Heart failure	
Yes (*n* = 379)	49.8 (10.3)
No (*n* = 1,830)	50.8 (10.0)
Stroke	
Yes (*n* = 190)	51.0 (10.1)
No (*n* = 2,019)	50.6 (10.1)
Hygiene—Toileting (*n* = 1,488[a])	
Independent (*n* = 71)	49.6 (11.9)
Setup or clean-up assistance (*n* = 77)	48.0 (8.3)
Supervision or touching assistance (*n* = 318)	50.0 (9.4)
Partial/moderate assistance (*n* = 357)	50.3 (9.9)
Substantial/maximal assistance (*n* = 331)	51.5 (9.9)
Dependent (*n* = 334)	52.8 (10.8)
Mobility—Lying to sitting (*n* = 1,848[a])	
Independent (*n* = 189)	49.3 (10.6)
Setup or clean-up assistance (*n* = 112)	50.6 (9.8)
Supervision or touching assistance (*n* = 526)	49.4 (9.3)
Partial/moderate assistance (*n* = 604)	50.6 (9.8)
Substantial/maximal assistance (*n* = 288)	51.8 (10.7)
Dependent (*n* = 129)	53.5 (10.3)

[a] Indicates significant ($p < 0.05$) association between patient/resident characteristic and PROMIS Anxiety T-score.

Based on the research literature, we generated several hypotheses or expectations for associations between the data elements and the patient/resident characteristics. We expected anxiety to be related to gender, level of assistance needed with toileting, and level of assistance needed in transferring from lying to sitting, such that patients/residents with higher PROMIS Anxiety T-score would tend to be female and have less independence in ADLs.[76]

Overall, the PROMIS Anxiety T-score is significantly associated with gender, age, disposition at discharge, sepsis, and the two ADLs. We review the statistical associations between variables in the bullets below.

Gender and Age

- Gender, overall, is significantly associated with the PROMIS Anxiety T-score ($t_{(2868)} = 3.43$, $p < 0.05$), such that females score higher (M = 51.0, SD = 10.0) than males (M = 49.7, SD = 10.1). A similar trend is observed at the setting level in HHAs but not in IRFs,

[76] Schoevers et al., 2003.

LTCHs, or SNFs. *This association conforms with our hypothesis and the research literature on the epidemiology of anxiety disorders in adults*[77] *and older adults.*[78]

- Age, overall, is significantly associated with PROMIS Anxiety T-score ($F_{(4,2854)} = 23.63$, $p < 0.01$), such that the three younger age groups (i.e., 18–44, 45–64, and 65–74) have significantly higher scores (M [SD] = 52.0 [11.3], 54.6 [12.3], and 51.5 [10.2], respectively) than the 90+ age group (M [SD] = 48.5 [9.1]; $t_{(2868)} = 2.04$, 7.54, and 4.57, respectively, $p < 0.05$). However, the 75–89 age group (M [SD] = 49.3 [9.6]) is not significantly different from the 90+ group. A similar trend is observed at the setting level (except in LTCHs, where there is no significant association), where younger age groups in general have higher PROMIS Anxiety T-scores. *This association conforms with our hypothesis and the research literature on the epidemiology of anxiety across the lifespan.*[79]

Length of Stay, Disposition at Discharge

- Length of stay, overall (M = 21.6 days, SD = 12.8), is not significantly associated with PROMIS Anxiety T-score (Pearson correlation $r = –0.02$, $p > 0.05$). At the setting level, length of stay is also not associated with the T-scores. *We did not have any expectation about this association.*
- Overall, disposition at discharge is significantly associated with PROMIS Anxiety T-score ($F_{(7,2813)} = 4.1$, $p < 0.001$). Relative to patients/residents discharged to all other placements, those discharged to LTCHs have the lowest T-score (M = 49.3, SD = 10.5) and those discharged to IRFs have the highest T-score (M = 53.3, SD = 11.5). At the setting level, however, disposition at discharge was not significantly associated with PROMIS Anxiety T-score. *We did not have any expectation about this association. These results are likely due to underlying patient/resident characteristics, such as age, gender, or clinical condition.*

Clinical Conditions

- There is an overall significant difference in PROMIS Anxiety T-scores ($t_{(2207)} = 3.94$, $p < 0.05$) between patients with sepsis (M = 53.8, SD = 11.2) and those without (M = 50.4, SD = 9.9). A similar trend is observed in LTCHs, where patients with sepsis have a higher T-score (M = 56.8, SD = 11.4) than those without (M = 53.0, SD = 10.3; $t_{(381)} = 2.67$, $p < 0.05$). In HHAs, however, the mean difference is reversed, such that those with sepsis have a significantly lower mean (M = 42.5, SD = 6.2) than those without (M = 49.5, SD = 9.4; $t_{(414)} = 2.09$, $p < 0.05$). There is no mean difference observed in IRFs or SNFs. *We did not predict an association between sepsis and anxiety. Although the experience of being treated for sepsis is likely stressful and potentially anxiety-producing on its own, there is some research to support the idea that sepsis is no different from other critical illness experiences with respect to its impact on a patients' mental health*[80]

[77] Kessler et al., 2005

[78] Schoevers et al., 2003.

[79] Kessler et al., 2012.

[80] Heyland et al., 2000.

and overall quality of life.[81] *We believe that the differences in anxiety T-scores between settings are likely related to underlying patient characteristics, such as gender, age, or acuity of overall clinical situation, and not related to the experience of sepsis itself.*

- There is no association of heart failure and stroke with PROMIS Anxiety T-scores overall or in any of the settings. *We did not have any expectations about these associations.*

ADLs: Toileting and Ability to Transfer from Lying to Sitting

- In the overall sample, level of assistance needed with toileting is significantly associated with PROMIS Anxiety T-score ($F_{(5,1482)} = 4.89$, $p < 0.001$), such that those needing minimal assistance (i.e., setup or clean-up assistance) have the lowest PROMIS Anxiety T-score (M = 48.0, SD = 8.3) as compared with those who need more assistance (e.g., for those with substantial/maximal assistance, M = 51.5, SD = 9.9) and those who are completely dependent (M = 52.8, SD = 10.8). However, those who are completely independent for toileting hygiene do not necessarily have lower T-scores (M = 49.6, SD = 11.9) as compared with those who need minimal assistance ($t_{(1482)} = 0.25$, $p > 0.05$). By setting, toileting hygiene is significantly associated with PROMIS Anxiety T-score only in IRFs, such that those who need more assistance in general have higher PROMIS Anxiety T-scores. *We hypothesized that patients/residents with less independence with toileting would have higher levels of anxiety, which we found to be generally true in the overall sample and in IRFs but not in other settings.*
- Overall ability to transfer from lying to sitting at different levels of need for assistance is significantly associated with PROMIS Anxiety T-score ($F_{(5,1842)} = 5.35$, $p < 0.001$), such that those who are independent have the lowest T-score (M = 49.3, SD = 10.6) as compared with those who need more assistance. By setting, ability to transfer from lying to sitting is only significantly associated with PROMIS Anxiety T-score in IRFs and LTCHs, such that those who need more assistance in general have higher PROMIS Anxiety T-scores. *As with toileting, this pattern of results was generally as expected and supports the validity of the PROMIS Anxiety T-score because it shows known associations with patient/resident characteristics.*

Time to Complete

Table 5.4 shows average time to complete the PROMIS Anxiety data elements overall and by setting for both versions. On average, the time to complete for version A (past 7 days) was 2.2 minutes (SD = 0.8) and ranged from 2.0 minutes (SD = 0.8) in HHAs to 2.2 minutes (SD = 0.8) in both LTCHs and SNFs. For version B (past 3 days), time to complete on average was also 2.2 minutes (SD = 0.9) and ranged from 1.9 minutes (SD = 0.8) in IRFs to 2.4 minutes (SD = 0.8) in HHAs. Time to complete was not significantly associated with setting type using version A ($F_{(3,754)} = 1.62$, $p > 0.05$). Time to complete was significantly associated with setting type using version B (past 3 days) ($F_{(3,800)} = 10.7$, $p < 0.01$), such that it took significantly more time to complete in HHAs than in IRFs ($t_{(800)} = 5.46$, $p < 0.01$) and LTCHs ($t_{(800)} = 3.3$, $p < 0.01$). It also took significantly more time to complete in SNFs than in IRFs ($t_{(800)} = 3.6$, $p < 0.01$).

[81] Granja et al., 2004.

Table 5.4. Time to Complete for PROMIS Anxiety Data Elements (minutes)

		HHA (*n* = 364)	IRF (*n* = 487)	LTCH (*n* = 267)	SNF (*n* = 444)	Overall (*n* = 1,562)
Past 7 days	Mean (SD)	2.0 (0.8)	2.1 (0.8)	2.2 (0.8)	2.2 (0.8)	2.2 (0.8)
Past 3 days	Mean (SD)	2.4 (0.8)	1.9 (0.8)	2.1 (0.8)	2.2 (0.9)	2.2 (0.9)
Combined	Mean (SD)	2.3 (0.8)	2.1 (0.8)	2.1 (0.8)	2.2 (0.9)	2.2 (0.8)

Time to complete was also evaluated according to urbanicity (urban versus nonurban), geographic region (Northeast, South, Midwest, West), facility ownership (for-profit versus nonprofit), and facility size (above or below setting-type median) to evaluate the generalizability of these performance results (see Tables A.34–A.37 in the appendix). No significant differences were found for time to complete the PROMIS Anxiety items in these sensitivity analyses.

Interrater Reliability

Table 5.5 shows kappa interrater reliability coefficients, with data from the two versions combined, overall and by setting. As described in more detail in Volume 3, paired assessment data for interrater reliability evaluation was collected on a subset of the National Beta Test communicative admission sample of patients/residents according to setting-level target totals. For example, each participating LTCH was asked to conduct 20 paired assessments to contribute to interrater reliability. Inclusion in interrater reliability data collection depended on paired facility staff and research nurse assessors' ability to schedule assessments. Kappas were computed on 925 patients/residents who were assessed by research nurse and facility/agency staff assessor pairs: 192 in HHAs, 249 in IRFs, 224 in LTCHs, and 260 in SNFs. Kappas for the PROMIS Anxiety items were excellent, ranging from 0.97 to 0.99. At the setting level, kappas were also excellent, ranging from 0.98 to 1.00 in HHAs, 0.95 to 0.99 in IRFs, 0.98 to 1.00 in LTCHs, and 0.96 to 0.99 in SNFs. There is little variation among kappas according to setting.

Table 5.5. Interrater Reliability Kappa or Weighted Kappa for PROMIS Anxiety Data Elements

Data Element	HHA (*n* = 192)	IRF (*n* = 249)	LTCH (*n* = 224)	SNF (*n* = 260)	Overall (*n* = 925)
How often past 3/7 days: felt worried (e3a)	1.00	0.98	0.98	0.98	0.98
How often past 3/7 days: overwhelmed by worries (e3b)	0.99	0.97	0.98	0.98	0.98
How often past 3/7 days: trouble paying attention (e3c)	0.99	0.96	0.98	0.96	0.97
How often past 3/7 days: felt nervous (e3d)	1.00	0.97	0.99	0.98	0.98
How often past 3/7 days: difficulty calming down (e3e)	0.98	0.95	0.98	0.97	0.97
How often past 3/7 days: focused on anxiety (e3f)	1.00	0.98	1.00	0.99	0.99
How often past 3/7 days: felt need for help with anxiety (e3g)	0.98	0.99	0.99	0.99	0.99
How often past 3/7 days: sudden feelings of panic (e3h)	1.00	0.97	1.00	0.98	0.98

Data Element	HHA (*n* = 192)	IRF (*n* = 249)	LTCH (*n* = 224)	SNF (*n* = 260)	Overall (*n* = 925)
Anxiety total[a]	1.00	0.99	1.00	0.99	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, 0.81–1.00 is excellent/almost perfect.
[a] Pearson correlation for total scores.

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.38–A.41 in the appendix). No significant differences were found for interrater reliability of the PROMIS Anxiety data elements in these sensitivity analyses.

Table 5.6 shows the percent agreement for the PROMIS Anxiety data elements overall and by setting. Overall, percent agreement was high for all data elements, ranging from 98 to 99 percent with minimal setting differences.

Table 5.6. Interrater Reliability—Percent Agreement for PROMIS Anxiety Data Elements

Data Element	HHA (*n* = 192)	IRF (*n* = 249)	LTCH (*n* = 224)	SNF (*n* = 260)	Overall (*n* = 925)
How often felt worried (e3a)	99	97	98	98	98
How often overwhelmed by worries (e3b)	99	98	99	98	98
How often had trouble paying attention (e3c)	99	97	98	97	98
How often felt nervous (e3d)	99	98	98	98	98
How often had difficulty calming down (e3e)	98	97	98	98	98
How often focused on anxiety (e3f)	100	100	100	99	99
How often felt need for help with anxiety (e3g)	99	100	98	99	99
How often sudden feelings of panic (e3h)	100	98	100	99	99

Scale Reliability

Because the PROMIS Anxiety data elements are typically interpreted at the scale level, we conducted additional analyses to evaluate the unidimensionality and internal-consistency reliability of these data elements. These analyses provided strong support for the scalability of the data elements. Data element total correlations ranged from 0.60 to 0.82, with an average inter–data element correlation of 0.55 and a Cronbach's alpha of 0.91. The explained common variance, which has a maximum of 1, is 0.81 for these data elements, indicating strong unidimensionality.

Admission to Discharge

Table 5.7 summarizes patterns of change from admission to discharge with data from both versions of PROMIS Anxiety data elements combined. As described in more detail in Volume 3, discharge data were collected on a subset of the National Beta Test admission sample of patients/residents. Availability of discharge data depended on advance notification of discharge and the ability to schedule assessments among the facility staff assessors in each participating site. Patterns are characterized as "no change" (scores stay the same at admission and discharge), "improve" (scores reflect improvement from admission to discharge), and "worsen" (scores indicate a decline from admission to discharge). Results are based on data from 776 patients/residents, 457 of whom received version A and 319 of whom received version B. Overall, responses demonstrated some degree of stability from admission to discharge, with an overall range of 47 to 75 percent experiencing no change. However, compared with admission, more patients/residents at discharge were showing improvement than were worsening in symptoms according to the PROMIS Anxiety data elements. For example, patients/residents were less likely to be "overwhelmed by worries" ($t_{(769)} = 7.81$, $p < 0.001$) at discharge relative to admission. Across all the data elements, the percentage worsening ranged from 7 to 17 percent, while the percentage showing improvement ranged from 18 to 36 percent. The majority of Anxiety total scores (56 percent) and T-scores (57 percent) also showed improvement from admission to discharge. There was little noticeable variation across settings.

Table 5.7. Admission to Discharge Results for PROMIS Anxiety Data Elements (percent)

Data Element	HHA (*n* = 145)	IRF (*n* = 328)	LTCH (*n* = 82)	SNF (*n* = 221)	Overall (*n* = 776)
How often felt worried (e3a)					
No change	48	48	39	47	47
Worsen	13	18	23	17	17
Improve	39	35	38	35	36
How often overwhelmed by worries (e3b)					
No change	67	65	51	71	66
Worsen	8	9	12	9	9
Improve	26	26	37	20	25
How often had trouble paying attention (e3c)					
No change	58	56	51	61	58
Worsen	13	11	18	10	12
Improve	29	33	30	29	31
How often felt nervous (e3d)					
No change	51	55	55	58	55
Worsen	13	10	18	12	12

Data Element	HHA (*n* = 145)	IRF (*n* = 328)	LTCH (*n* = 82)	SNF (*n* = 221)	Overall (*n* = 776)
Improve	36	35	27	30	33
How often had difficulty calming down (e3e)					
No change	70	65	54	72	67
Worsen	5	7	12	8	7
Improve	25	29	33	21	26
How often focused on anxiety (e3f)					
No change	68	65	60	76	68
Worsen	9	8	18	5	8
Improve	23	27	22	19	23
How often felt need for help with anxiety (e3g)					
No change	71	64	67	72	68
Worsen	6	9	10	10	9
Improve	23	27	23	18	24
How often sudden feelings of panic (e3h)					
No change	79	72	63	80	75
Worsen	3	8	13	5	7
Improve	18	20	23	14	18
Anxiety total score					
No change	24	23	17	30	25
Worsen	17	19	29	19	20
Improve	59	58	54	50	56
Anxiety T-score					
No change	22	19	9	24	20
Worsen	16	22	35	24	23
Improve	62	59	56	52	57

The summed score to T-score conversion table for the eight PROMIS Anxiety data elements administered in the National Beta Test is included as Table A.42 in the appendix.

Assessor Feedback

Facility staff indicated on the assessor survey that PROMIS Anxiety was somewhat to moderately clinically useful, although ratings from both facility staff and research nurses were lower than for other data elements. Survey participants deemed PROMIS Anxiety to be slightly to somewhat difficult to collect, with facility staff ratings falling in the high end of the distribution for burden and research nurse ratings falling in the middle to high end of the distribution for burden, relative to all other data elements. Respondents in both groups also rated PROMIS Anxiety relatively higher than other data elements for burden to the patient/resident.

Similar to the PROMIS Depression items, when survey participants were asked about factors impacting data collection, they noted patient discomfort and fatigue as prevalent issues.

When discussed in focus groups, facility staff felt that a strength of these questions, as a whole, was utility for identifying signs and symptoms of anxiety among patients/residents who would feel stigmatized by expressly labeling anxiety through these questions, similar to the PROMIS Depression data elements. These questions were more palatable to patients/residents than both the PHQ and the PROMIS Depression questions. Field staff also stated in the focus groups that assessment of anxiety was clinically relevant to them, especially for transfers from one setting to another because they perceived that mood affects overall health.

However, not all members of these facility staff focus groups agreed with the positive feedback mentioned above, as evidenced by, on average, facility staff reporting a relatively high clinical burden for these data elements in the assessor survey. In focus groups, some facility staff thought that it was burdensome and not clinically useful to ask all the questions. These individuals felt that some questions were redundant and said they would have preferred a shorter version or one with skip patterns. Further, in some PAC settings the question about feeling worried was not helpful for care planning:

> And I do feel like on the anxiety part, the "I felt worried," . . . we should all be worried if we're in rehab, because we really don't know how it's going to turn out, you know what I mean, realistically.
>
> —Durham, SNF Staff

Similar to feedback for the PROMIS Depression questions, facility staff in focus groups questioned the validity of the patient/resident answers because some patients seemed to give answers that were disingenuous or inconsistent. If a spouse or family member was present, these individuals might contradict the patient/resident, leaving the assessor wondering whether the patient/resident was just giving an answer that she or he thought the assessor wanted to hear.

In summary, both the assessor survey and focus groups indicated that, similar to the PHQ and PROMIS Depression, standardizing anxiety assessments was of clinical value to the assessors, but assessors had mixed feedback on PROMIS Anxiety data elements given the burden it posed on patients/residents. Further, focus groups identified concerns about the validity of these data elements among the facility staff.

Summary

Results for PROMIS Anxiety data elements indicate moderate overall support for cross-setting standardization. Assessors considered PROMIS Anxiety data elements to be somewhat to moderately clinically useful but rated it as having relatively high data collection burden and burden to the patient/resident. Although some assessors had concerns that the PROMIS Anxiety data elements were redundant and questioned the validity of patient/resident responses, assessors also noted that the PROMIS Anxiety data elements were more palatable to patients/residents

than the PHQ-2 to 9 and PROMIS Depression. We also found that the data elements performed similarly according to setting, gender, and age (i.e., did not observe any DIF), which supports their use in the PAC population. For interrater reliability, kappas were excellent and percent agreement was high, with little variation across the two versions (past 7 days versus past 3 days) or at the setting level. Responses demonstrated some degree of stability from admission to discharge, with patients/residents more likely to show improvement in symptoms than to worsen at discharge. The combined results for PROMIS Anxiety data elements are somewhat mixed, showing excellent interrater reliability but only moderate feasibility because of the relatively high burden, and some to moderate clinical utility as a candidate data element for standardization across PAC settings.

6. Pain Interview

Data Element Description

The Pain Interview includes data elements that assess the presence of pain and pain frequency, along with three data elements evaluating for pain interference (with sleep, therapy activities, and other activities), pain severity, and relief from pain due to pain treatments or medications. These data elements are completed through patient/resident interview. Similar versions of some of the Pain Interview data elements are included in the OASIS-D and the MDS. In the National Beta Test, the Pain Interview data elements were collected in two versions. One version, shown in Figure 6.1, asks about pain experiences in "the past 3 days." The second version was identical except that it referenced "the past 5 days" to align more closely with the Pain Interview currently used in the MDS. In addition, the Pain Interview was included in the repeat assessment evaluation and therefore was evaluated repeatedly on the same patient/resident by the same assessor on Days 3, 5, and 7, per the repeat assessment design.

Figure 6.1. Pain Interview Data Elements

D1. ASK PATIENT/RESIDENT:
"Have you had pain or hurting any time in the last 5 days?"

- ☐ 0 = No **[SKIP to D-TIME]**
- ☐ 1 = Yes
- ☐ 7 = **Patient/resident declined to respond [SKIP to D-TIME]**
- ☐ 9 = **Unable to answer or no response [SKIP to D-TIME]**

D2. ASK PATIENT/RESIDENT:
"How much of the time have you experienced pain or hurting over the last 5 days?

- ☐ 1 = Rarely or not at all
- ☐ 2 = Occasionally
- ☐ 3 = Frequently
- ☐ 4 = Almost Constantly
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unable to answer or no response**

D3. ASK PATIENT/RESIDENT:
"Over the past 5 days, how much of the time has pain made it hard for you to sleep?"

- ☐ 1 = Rarely or not at all

- ☐ 2 = Occasionally
- ☐ 3 = Frequently
- ☐ 4 = Almost Constantly
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unable to answer or no response**

D4a. ASK PATIENT/RESIDENT:
"Over the past 5 days, have you been offered any rehabilitation therapies (e.g., physical therapy, occupational therapy, speech therapy) by your care providers?"

- ☐ 0 = No [**SKIP to D4c**]
- ☐ 1 = Yes
- ☐ 7 = **Patient/resident declined to respond** [**SKIP to D4c**]
- ☐ 9 = **Unable to answer or no response** [**SKIP to D4c**]

D4b. ASK PATIENT/RESIDENT:
"Over the past 5 days, how often have you limited your participation in rehabilitation therapy sessions due to pain?"

- ☐ 1 = Rarely or not at all
- ☐ 2 = Occasionally
- ☐ 3 = Frequently
- ☐ 4 = Almost Constantly
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unable to answer or no response**

D4c. ASK PATIENT/RESIDENT:
"Over the past 5 days, how much of the time have you limited your day-to-day activities (excluding rehabilitation therapy sessions) because of pain?"

- ☐ 1 = Rarely or not at all
- ☐ 2 = Occasionally
- ☐ 3 = Frequently
- ☐ 4 = Almost Constantly
- ☐ 7 = **Patient/resident declined to respond**
- ☐ 9 = **Unable to answer or no response**

D5. SAY TO PATIENT/RESIDENT:
"Please rate your worst pain over the last 5 days on a zero to ten scale, with zero being no pain and ten as the worst pain you can imagine."

- ☐ 0 = No pain
- ☐ 1
- ☐ 2

☐ 3
☐ 4
☐ 5
☐ 6
☐ 7
☐ 8
☐ 9
☐ 10 = Worst pain imaginable
☐ 77 = **Patient/resident declined to respond**
☐ 99 = **Unknown or unable to assess**

D6. ASK PATIENT/RESIDENT:
"Over the past 5 days how much relief have you felt from pain due to pain treatments and/or medications?"

☐ 1 = No relief
☐ 2 = Some relief
☐ 3 = Quite a bit of relief
☐ 4 = Very much relief
☐ 8 = Not applicable- patient/resident has not received pain treatments or medications in the past 5 days
☐ 7 = **Patient/resident declined to respond**
☐ 9 = **Unable to answer or no response**

Testing Objectives

As stated above, two versions of the Pain Interview data elements were administered during the National Beta Test: a version referencing pain in the past 3 days, and another referencing pain in the past 5 days. The 3-day version was administered in Market Group A, and the 5-day version in Market Group B. For more details on Market Group A and B samples and characteristics, please refer to Volume 3. Basic descriptive statistics (e.g., frequencies, means, SDs) are first presented for admission data to characterize the responses to the Pain Interview data elements according to version for patients/residents in each setting and for the overall sample. Next, we present differences in the versions that we evaluated by conducting t-tests to compare data element and scale level means across the two versions; a Cohen's *d* value of greater than 0.2 was used to indicate meaningful version differences. Because we saw little evidence of performance differences according to version, the majority of results, with the exception of feasibility results (item frequencies, rates of missingness, and time to complete), are presented for the two versions combined. Specifically, we used the combined admission response data to (1) examine Pain Interview data elements by patient/resident characteristics and clinical groups of interest (i.e., known groups validity), (2) evaluate interrater reliability (kappa and

percent agreement), and (3) compare frequencies based on repeat assessments at Days 3, 5, and 7 and from admission to discharge to inform stability or possible change over time. Regarding this third point, in addition to being administered at both admission and discharge, the Pain Interview was administered repeatedly on the same patient/resident by the same assessor on Days 3, 5, and 7. Therefore, an additional objective was to understand whether there were significant and meaningful changes in rates or scores depending on the day a patient/resident was assessed.

Results

Feasibility

Frequencies/Missing

Tables 6.1 and 6.2 show the percentage of responses at admission for each Pain Interview data element, both overall and by setting. Frequencies are presented for both market versions (A: "past 3 days" and B: "past 5 days") side-by-side for ease of comparison. Version A ("past 3 days") was administered to 1,520 patients/residents: 229 in HHAs, 488 in IRFs, 256 in LTCHs, and 547 in SNFs. Version B ("past 5 days") was administered to 1,511 patients/residents: 415 in HHAs, 293 in IRFs, 228 in LTCHs, and 575 in SNFs. Overall, just above 97 percent of the sample was administered one of the two versions. Among these patients/residents, missing data at the data element level did not exceed 2.8 percent for either version, with minimal setting differences.

Overall, a majority of patients/residents indicated pain presence, with similar rates between versions (version A: 75 percent, version B: 80 percent), and these patients/residents were consequently administered the remaining Pain Interview data elements (those who reported no pain skipped out of the Pain Interview data element). Although respondents with pain presence did not often report pain frequency as "rarely or not at all," they were about equally likely to endorse one of the other three response categories. Endorsement of response categories for the three pain interference data elements (sleep, therapy activities, other activities) was much less likely to be in the "almost constantly" category, tending to be less frequent, particularly for the activities data elements. Patients'/residents' worst pain rating on the 0–10 scale averaged just over 7 overall; nearly 95 percent of those experiencing pain during the period indicated they were getting at least some relief from the pain due to treatment and/or therapy, and more than half of respondents said they had experienced "quite a bit" (35 percent) of relief or "very much" (17 percent) relief.

Table 6.1. Overall and Setting-Specific Response Frequencies for Pain Interview Data Elements: Market Group A (percent)

Data Element	HHA (*n* = 229)	IRF (*n* = 488)	LTCH (*n* = 256)	SNF (*n* = 547)	Overall (*n* = 1,520)
Any pain (d1)					
Yes	68	82	73	73	75
Among those who experienced any pain					
Number of patients/residents	156	398	188	398	1,140
How often experienced pain (d2)					
Rarely or not at all	8	6	7	6	6
Occasionally	31	28	19	29	27
Frequently	37	41	33	38	38
Almost constantly	24	24	41	26	28
How often pain made it hard to sleep (d3)					
Rarely or not at all	38	30	25	40	33
Occasionally	31	29	19	24	26
Frequently	20	29	35	24	27
Almost constantly	11	13	21	13	14
Offered rehab therapies (d4a)					
Yes	84	98	74	92	90
Yes N	131	389	138	365	1,023
How often limited rehab due to pain (d4b)					
Rarely or not at all	75	75	70	74	74
Occasionally	15	17	19	14	16
Frequently	6	5	7	10	7
Almost constantly	4	3	4	2	3
How often limited daily activities due to pain (d4c)					
Rarely or not at all	37	56	42	33	43
Occasionally	31	17	17	29	23
Frequently	16	18	17	24	19
Almost constantly	17	9	24	14	14
Rate worst pain (d5) Mean (SD)	6.3 (2.4)	7.3 (2.5)	7.6 (2.3)	7.1 (2.5)	7.2 (2.5)
How much pain relief due to pain treatments and/or medications					
None	3	6	8	5	5
Some	50	39	48	46	44
Quite a bit	34	38	27	28	32
Very much	13	17	17	21	18

Table 6.2. Overall and Setting-Specific Response Frequencies for Pain Interview Data Elements: Market Group B (percent)

Data Element	HHA (*n* = 415)	IRF (*n* = 293)	LTCH (*n* = 228)	SNF (*n* = 575)	Overall (*n* = 1,511)
Any pain (d1)					
Yes	80	75	82	82	80
Among those who experienced any pain					
Number of patients/residents	333	220	187	474	1,214
How often experienced pain (d2)					
Rarely or not at all	7	7	9	7	7
Occasionally	34	30	29	32	32
Frequently	31	39	35	36	35
Almost constantly	28	24	28	25	26
How often pain made it hard to sleep (d3)					
Rarely or not at all	41	35	34	34	36
Occasionally	28	30	29	31	30
Frequently	18	21	23	22	21
Almost constantly	13	13	14	13	13
Offered rehab therapies (d4a)					
Yes	75	99	88	93	88
Yes N	248	217	164	438	1,067
How often limited rehab due to pain (d4b)					
Rarely or not at all	74	77	55	72	71
Occasionally	14	17	16	18	16
Frequently	7	5	19	7	8
Almost constantly	5	1	10	3	4
How often limited daily activities due to pain (d4c)					
Rarely or not at all	42	52	43	47	46
Occasionally	24	20	20	23	22
Frequently	18	14	23	20	19
Almost constantly	16	14	14	10	13
Rate worst pain (d5) Mean (SD)	6.7 (2.7)	7.3 (2.6)	7.4 (2.5)	7.2 (2.6)	7.1 (2.6)
How much pain relief due to pain treatments and/or medications					
None	5	4	5	7	6
Some	46	41	46	38	42
Quite a bit	33	33	33	37	35
Very much	16	22	16	17	17

Results of the comparison of data element means by version are shown in Table A.43 in the appendix. Although three data elements (pain presence, pain frequency, pain effect on sleep) out of seven showed statistically significant differences by version, these were well below the Cohen's *d* value of 0.2.

Known Groups Validity

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

Table 6.3 shows the percentage of patients/residents reporting any pain (based on the Pain Presence data element) and the percentage of those patients/residents who said that pain interfered with their sleep either "frequently" or "almost constantly." The table shows these respondents 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 (18–44, 45–64, 65–74, 75–89, or 90 and over), length of stay (in days), disposition at discharge (e.g., to another PAC setting, home, to hospital), sepsis, heart failure, stroke, and two ADLs: toileting (not available for HHA patients) and ability to transfer from lying to sitting. As a reminder, these clinical conditions were chosen based on their common occurrence across settings, their frequent relationship with many of the data elements tested in the National Beta Test, and their availability in all four settings (i.e., equivalent information was collected on the OASIS, IRF-PAI, LCDS, and MDS). Setting-specific results for these individual characteristics are presented in the appendix in Tables A.44–A.47.

Table 6.3. Frequencies for Pain Presence and Frequent or Almost Constant Pain Interference with Sleep, Therapy, and Other Activities by Patient/Resident Characteristics and Clinical Groups (percent)

Patient/Resident Characteristics and Clinical Groups	Any Pain	Interference with Sleep
Gender (*n_any* = 2,926[a]; *n_int* = 2,270)		
Male (*n_any* = 1,204; *n_int* = 898)	74.5	36.6
Female (*n_any* = 1,722; *n_int* = 1,372)	79.7	37.2
Age (*n_any* = 2,915[a]; *n_int* = 2,261[a])		
18–44 (*n_any* = 42; *n_int* = 37)	88.1	51.4
45–64 (*n_any* = 308; *n_int* = 263)	85.4	56.5
65–74 (*n_any* = 915; *n_int* = 767)	83.8	41.6
75–89 (*n_any* = 1,337; *n_int* = 993)	74.3	30.5

Patient/Resident Characteristics and Clinical Groups	Any Pain	Interference with Sleep
90+ (*n_any* = 313; *n_int* = 201)	64.2	23.7
Length of stay (*n_any* = 2,580; *n_int* = 1,999; mean, SD)	No: 21.3 (13.0) Yes: 21.6 (12.7)	No: 21.7 (13.0) Yes: 21.4 (12.1)
Disposition at discharge (*n_any* = 2,876[a]; *n_int* = 2,234[a])		
Home (*n_any* = 1,343; *n_int* = 1,023)	76.1	35.1
Hospital (*n_any* = 200; *n_int* = 162)	81.0	40.7
Hospice (*n_any* = 40; *n_int* = 28)	70.0	32.1
HHA (*n_any* = 623; *n_int* = 515)	82.7	37.5
IRF (*n_any* = 51; *n_int* = 41)	80.4	51.2
LTCH (*n_any* = 13; *n_int* = 8)	61.5	37.5
SNF (*n_any* = 282; *n_int* = 219)	77.7	41.9
Other (*n_any* = 324; *n_int* = 238)	73.5	33.8
Clinical conditions (*n_any* = 2,257; *n_int* = 1,759)		
Sepsis		
Yes (*n_any* = 152; *n_int* = 117[a])	77.0	49.6
No (*n_any* = 2,105; *n_int* = 1,642)	78.0	37.1
Heart failure		
Yes (*n_any* = 385[a]; *n_int* = 278)	72.2	33.6
No (*n_any* = 1,872; *n_int* = 1,481)	79.1	38.8
Stroke		
Yes (*n_any* = 197[a]; *n_int* = 128)	65.0	32.3
No (*n_any* = 2,060; *n_int* = 1,631)	79.2	38.4
Hygiene—Toileting (*n_any* = 1,525[a]; *n_int* = 1,183[a])		
Independent (*n_any* = 73; *n_int* = 58)	79.5	46.6
Setup or clean-up assistance (*n_any* = 77; *n_int* = 63)	81.8	33.3
Supervision or touching (*n_any* = 321; *n_int* = 231)	72.0	32.6
Partial/moderate (*n_any* = 367; *n_int* = 276)	75.2	35.3
Substantial/maximal (*n_any* = 336; *n_int* = 267)	79.5	44.0
Dependent (*n_any* = 351; *n_int* = 288)	82.1	44.2
Mobility—Lying to sitting (*n_any* = 1,884[a]; *n_int* = 1,457[a])		
Independent (*n_any* = 193; *n_int* = 142)	73.6	43.0
Setup or clean-up (*n_any* = 114; *n_int* = 82)	71.9	35.8
Supervision or touching (*n_any* = 528; *n_int* = 389)	73.7	31.4
Partial/moderate (*n_any* = 619; *n_int* = 491)	79.3	38.3
Substantial/maximal (*n_any* = 293; *n_int* = 241)	82.3	41.4
Dependent (*n_any* = 137; *n_int* = 112)	81.8	46.0

NOTES: *n_any* = number of patients/residents who reported the presence of any pain; *n_int* = number of patients/residents who said that pain interfered with their sleep "frequently" or "almost constantly."
[a] Significant ($p < 0.05$) associations with interference due to pain.

Based on the research literature, we generated several hypotheses or expectations for associations between the Pain Presence data element and patient/resident characteristics. A recent study of long-stay nursing home residents found more documented pain in females than males and in younger persons than older persons (highest in 18–64 age group and lowest in 85 and older).[82] An earlier study found more pain among those with less independence in ADLs.[83] Therefore, we expected reports of pain to be related to gender, age, and toileting and mobility (ability to transfer from lying to sitting) ADL impairments. We did not have any hypotheses or expectations for the relationship between these characteristics and conditions and whether pain interferes with sleep.

Significant effects involving one or both of the pain data elements examined were observed for all characteristics either based on the full sample combined across settings (i.e., as tabulated below) and/or in one of the specific settings (i.e., as tabulated in Tables A.44–A.47 in the appendix). We review the statistical associations between variables in the bullets below.

Gender and Age

- Gender, overall, was significantly associated with pain presence ($\chi^2_{(1)} = 10.55$, $p < 0.01$), such that a greater percentage of females (79.7 percent) reported pain presence compared with males (74.5 percent). Similar trends for pain presence and gender were found in all settings but only reached statistical significance in the IRF setting ($\chi^2_{(1)} = 8.86$, $p < 0.01$). Gender was related to pain interference with sleep in HHAs, where women (33.9 percent) were more likely than men (24.4 percent) to report frequent or almost constant pain interference with sleep ($\chi^2_{(1)} = 4.48$, $p < 0.05$), but this trend was not observed or did not reach significance in other settings and in the overall sample. *These findings are consistent with our expectations to observe more reports of pain in women than in men. We did not have any expectations regarding gender and sleep interference; however, other than in HHAs, it appears that pain affects men's and women's sleep equally.*
- Age was significantly associated with pain presence overall ($\chi^2_{(4)} = 74.50$, $p < 0.01$) and in all settings except IRFs. Overall, patients/residents in the younger age groups tended to report experiencing pain at higher rates than those in the older age groups. For example, across all respondents, 88.1 percent of patients/residents aged 18–44 reported pain presence, whereas the rate was only 64.2 percent for those aged 90 and over. A similar trend emerged in HHAs ($\chi^2_{(4)} = 29.15$, $p < 0.001$), LTCHs ($\chi^2_{(4)} = 27.12$, $p < 0.001$), and SNFs ($\chi^2_{(4)} = 42.84$, $p < 0.001$). Older patients/residents were also less likely than those in the younger categories to report pain interfering with sleep frequently or almost constantly. For example, 56.5 percent of 45–64-year-olds reported interference with sleep, whereas only 30.5 percent of those in the 75–89 age range and 23.7 percent of those 90 and older did so ($\chi^2_{(4)} = 85.72$, $p < 0.001$). A similar trend emerged in HHAs ($\chi^2_{(4)} = 21.71$, $p < 0.001$), IRFs ($\chi^2_{(4)} = 27.01$, $p < 0.001$), and SNFs ($\chi^2_{(4)} = 42.42$, $p <$

[82] Hunnicutt et al., 2017.

[83] Won et al., 1999.

0.001) but not in LTCHs. *The relationship between age and pain presence is consistent with a recent study that found younger long-term nursing-home residents report pain at higher rates.*[84] *Although we did not form hypotheses related to age and pain interfering with sleep, findings are consistent with the idea that younger patients/residents both experience and are more bothered by pain than older patients/residents, likely due at least in part to the underlying clinical conditions for which they are receiving PAC services.*

Length of Stay, Disposition at Discharge

- We observed no associations of pain presence or interference with sleep with length of stay for the overall sample; patients/residents who reported pain had similar lengths of stay to patients/residents without pain, and patients/residents who reported high levels of sleep interference due to pain had similar lengths of stay to patients/residents without sleep interference due to pain. However, length of stay was significantly longer for LTCH patients who reported pain (M = 24.8 days, SD = 11.3) relative to those who did not (M = 20.0 days, SD = 9.3, $t_{(407)} = 3.63$, $p < 0.001$). *We did not generate hypotheses related to length of stay, and generally found no associations. However, the association we did observe in LTCH patients is likely due to the underlying clinical conditions for which these patients are receiving care. For example, patients with serious illness or recovering from serious trauma in an LTCH (i.e., patients with longer stays) may unfortunately also be experiencing the type of pain that interferes with their ability to sleep.*
- Overall, disposition at discharge was significantly associated with pain presence, ($\chi^2_{(7)} = 18.82$, $p < 0.01$). Rates of pain presence were lowest among patients/residents discharged to LTCHs (61.5 percent) and hospice (70.0 percent), and highest among patients/residents discharged to HHAs (82.7 percent), hospitals (81.0 percent), and IRFs (80.4 percent). At the setting level, disposition at discharge was also associated with pain presence in LTCHs ($\chi^2_{(7)} = 21.25$, $p < 0.01$) and SNFs ($\chi^2_{(7)} = 20.02$, $p < 0.01$) but not in HHAs or IRFs. There were no associations of pain effect on sleep with disposition at discharge overall or at the setting level. *The association between reports of pain and disposition at discharge is difficult to interpret. We had no expectations about this association, and the findings are likely related to the underlying patient/resident clinical conditions. Also, because the Pain Presence data element does not assess the frequency or intensity of pain, it is not appropriate to interpret these associations for patients/residents with more or less pain. Rather, these patterns are true when thinking about patients/residents who have any pain, versus those who report no pain.*

Clinical Conditions

- For patients/residents with sepsis, there was no association with pain presence in the overall sample, nor were there associations in IRFs, LTCHs, and SNFs. However, HHA patients with sepsis were much *less* likely to report pain presence (44.4 percent) relative to HHA patients who did not have sepsis (76.9 percent; $\chi^2_{(1)} = 5.11$, $p < 0.05$). Overall, there was a significant association of sepsis with pain interference with sleep ($\chi^2_{(1)} =$

[84] Hunnicutt et al., 2017.

7.18, $p < 0.01$). Those with sepsis were more likely to report interference with sleep due to pain (49.6 percent vs. 37.1 percent). A similar effect was noted among IRF patients (59.1 percent vs. 36.3 percent ; $\chi^2_{(1)} = 4.64$, $p < 0.05$). However, this association did not emerge in HHAs, LTCHs, or SNFs. *Although sepsis is a severe medical condition, we did not observe an association with higher levels of pain in the overall sample, and in fact HHA patients with sepsis reported pain at lower rates, suggesting perhaps that their conditions, including any pain, are being adequately managed, compared with HHA patients without sepsis, who may have clinical conditions for which some level of pain is normative (e.g., recovery from joint replacement, osteoarthritis not related to PAC episode). The association with pain interfering with sleep in patients/residents with sepsis is interesting and likely due to the underlying experience of patients/residents receiving treatment for sepsis and related conditions (e.g., organ failure, complications of surgery).*

- Those with heart failure ($\chi^2_{(1)} = 8.86$, $p < 0.01$) and stroke ($\chi^2_{(1)} = 21.08$, $p < 0.001$) were significantly less likely than those without to report pain presence overall (72.2 percent versus 79.1 percent and 65.0 percent versus 79.2 percent, respectively), and similar trends were noted among IRF patients for both heart failure ($\chi^2_{(1)} = 3.89$, $p < 0.05$) and stroke ($\chi^2_{(1)} = 45.21$, $p < 0.001$). Pain interference with sleep was reported at lower rates for HHA patients with heart failure ($\chi^2_{(1)} = 4.89$, $p < 0.05$) and at lower rates for IRF patients suffering from stroke ($\chi^2_{(1)} = 4.25$, $p < 0.05$), although pain interference with sleep was not associated with these conditions in the overall sample or in other settings. *We had no expectations about the relationship between pain and pain interference with sleep with stroke and heart failure. Indication of lower rates of pain and pain interference in patients/residents with these conditions in some settings suggests that these are relatively lower-pain conditions when compared with other reasons that patients/residents are receiving PAC services.*

ADLs: Toileting and Ability to Transfer from Lying to Sitting

- The association of toileting independence with pain presence and pain effect on sleep was observed only for the overall combined sample. Specifically, for pain presence, those in the middle independence category (supervision or touching) had lower rates (72.0 percent) relative to all the other independence categories (range 75.2–82.1 percent ; $\chi^2_{(5)} = 12.67$, $p < 0.05$). For pain interference with sleep, the completely independent group (46.6 percent) and the substantial/maximal (44.0 percent) and dependent (44.2 percent) groups had higher interference with sleep than those in the middle three independence categories (range 32.6 percent to 35.3 percent; $\chi^2_{(5)} = 13.73$, $p < 0.05$). *We expected to find an association between the pain data elements and level of dependence on ADLs, which we found. However, the direction of the association is more complex than we predicted. Rather than only patients/residents who need more assistance with toileting reporting higher levels of pain and pain that interferes with sleep, patients/residents who were more independent also had slightly higher rates. This U-shaped relationship is likely related to the underlying and heterogeneous clinical conditions of the patients/residents who are receiving PAC care.*
- For ability to transfer from lying to sitting, pain presence tended to be reported at higher rates for those with higher dependence in the overall sample ($\chi^2_{(5)} = 14.45$, $p < 0.05$), as well as in IRFs ($\chi^2_{(5)} = 21.20$, $p < 0.001$). This trend was generally evident as well for

pain interference with sleep ($\chi^2_{(5)} = 12.92$, $p < 0.05$) in the overall sample and among LTCH patients ($\chi^2_{(5)} = 14.26$, $p < 0.05$), although rates for the completely independent group do not quite fit the pattern because they are somewhat higher than would be expected. *This finding is similar to the associations observed for the toileting ADL and suggests that patients/residents who are the most dependent and the most independent report higher rates of pain and pain interfering with sleep than patients/residents who are "in the middle" on ADL abilities. This likely reflects the diversity of conditions and patient/resident characteristics found within PAC settings, ranging from otherwise healthy patients/residents who are recovering from accidents or joint replacement, to patients/residents who may be living with serious chronic conditions and are completing a PAC stay as a step-down from an acute hospitalization.*

Time to Complete

Table 6.4 shows the average time to complete Pain Interview data elements overall and by setting. Time to complete is presented for each market version separately and combined. Moreover, time to complete is presented for three groups: all patients/residents who completed pain data elements ("all"), only those who reported pain ("pain"), and only those who reported no pain ("no pain").

For all patients/residents who completed the Pain Interview data elements, overall time to complete was not statistically different between version A and B, taking, on average, 2.6 minutes to complete. For version A, time to complete was not associated with setting type. For version B, time to complete was significantly associated with setting type ($F_{(3,863)} = 12.78$, $p < 0.01$), such that it took significantly less time in IRFs than in any other setting [HHAs: $t_{(436)} = 5.44$, $p < 0.01$; LTCHs: $t_{(348)} = 5.75$, $p < 0.01$; SNFs: $t_{(433)} = 4.65$, $p < 0.01$]. For both versions combined, time was associated with setting, $F_{(3,1773)} = 4.32$, $p < 0.01$; however, after adjusting for multiple comparisons, there were no significant differences among settings.

For patients/residents reporting no pain ($n = 440$) the average time to complete was 1.3 minutes (SD = 1.0) for those asked in the past 3 days, 1.0 minutes (SD = 0.6) to complete for those asked in the past 5 days, and 1.2 minutes (SD = 0.9) to complete for the combined population. Setting-specific time ranged from 1.0 minutes (SD = 0.5) in IRFs and 1.4 minutes (SD = 1.5) in HHAs for those asked in the past 3 days, 0.8 minutes (SD = 0.4) in IRFs and 1.2 minutes (SD = 0.5) in HHAs for those asked in the past 5 days, and 0.9 minutes (SD = 0.5) in IRFs and 1.3 minutes (SD = 1.1) in HHAs for the combined population. For patients/residents reporting pain ($n = 1{,}331$), the average time to complete was 3.1 minutes (SD = 1.3) for those asked in the past 3 days, 3.0 minutes (SD = 1.2) to complete for those asked in the past 5 days, and 3.0 minutes (SD = 1.3) to complete for the combined population. Setting-specific time ranged from 2.9 minutes (SD = 1.3) in SNFs and 3.3 minutes (SD = 1.4) in HHAs and LTCHs for those asked in the past 3 days, 2.5 minutes (SD = 1.1) in IRFs and 3.2 minutes (SD = 1.3) in LTCHs for those asked in the past 5 days, and 2.9 minutes (SD = 1.2) in IRFs and 3.2 minutes (SD = 1.3) in HHAs and LTCHs for the combined population.

Table 6.4. Time to Complete Pain Data Elements

		HHA	IRF	LTCH	SNF	Overall
All	Number of patients/residents	440	533	321	483	1,777
Past 3 days	Mean (SD)	2.7 (1.7)	2.7 (1.3)	2.7 (1.6)	2.4 (1.4)	2.6 (1.5)
Past 5 days	Mean (SD)	2.7 (1.4)	2.0 (1.2)	2.8 (1.4)	2.6 (1.4)	2.6 (1.4)
Combined	Mean (SD)	2.7 (1.5)	2.4 (1.3)	2.7 (1.5)	2.5 (1.4)	2.6 (1.4)
No Pain	Number of patients/residents	122	120	81	117	440
Past 3 days	Mean (SD)	1.4 (1.5)	1.0 (0.5)	1.4 (1.0)	1.3 (1.0)	1.3 (1.0)
Past 5 days	Mean (SD)	1.2 (0.5)	0.8 (0.4)	1.3 (0.9)	0.9 (0.5)	1.0 (0.6)
Combined	Mean (SD)	1.3 (1.1)	0.9 (0.5)	1.3 (0.9)	1.2 (0.9)	1.2 (0.9)
Pain	Number of patients/residents	318	411	238	364	1,331
Past 3 days	Mean (SD)	3.3 (1.4)	3.1 (1.1)	3.3 (1.4)	2.9 (1.3)	3.1 (1.3)
Past 5 days	Mean (SD)	3.2 (1.2)	2.5 (1.1)	3.2 (1.3)	3.0 (1.2)	3.0 (1.2)
Combined	Mean (SD)	3.2 (1.3)	2.9 (1.2)	3.2 (1.3)	3.0 (1.2)	3.0 (1.3)

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.48–A.51 in the appendix). No significant differences were found for time to complete Pain data elements in these sensitivity analyses.

Interrater Reliability

Table 6.5 shows kappa interrater reliability coefficients for the two versions combined overall and by setting. As described in more detail in Volume 3, paired assessment data for interrater reliability evaluation was collected on a subset of the National Beta Test communicative admission sample of patients/residents according to setting-level target totals. For example, each participating LTCH was asked to conduct 20 paired assessments to contribute to interrater reliability. Inclusion in interrater reliability data collection depended on paired facility staff and research nurse assessors' ability to schedule assessments. Kappas were computed on 953 patients/residents who were assessed by research nurse and facility/agency staff assessor pairs: 197 in HHAs, 256 in IRFs, 232 in LTCHs, and 268 in SNFs. Overall, kappas for Pain Interview data elements were excellent, ranging from 0.96 to 0.98. At the setting level, kappas were also excellent, ranging from 0.95 to 0.99 in HHAs, 0.95 to 1.00 in IRFs, 0.96 to 1.00 in LTCHs, and 0.83 to 0.99 in SNFs. Generally speaking, kappas were all excellent with minimal setting differences.

Table 6.5. Interrater Reliability Kappa or Weighted Kappa for Pain Interview Data Elements

Data Element	HHA (*n* = 197)	IRF (*n* = 256)	LTCH (*n* = 232)	SNF (*n* = 268)	Overall (*n* = 953)
Any pain (d1)	0.99	0.97	1.00	0.99	0.98
How often experienced pain (d2)	0.97	0.99	0.96	0.97	0.97
How often pain made it hard to sleep (d3)	0.96	0.98	0.98	0.99	0.98
Offered rehab therapies (d4a)	0.96	1.00	1.00	0.83	0.96
How often limited rehab due to pain (d4b)	0.95	0.96	0.98	0.97	0.97
How often limited daily activities due to pain (d4c)	0.97	0.98	0.99	0.98	0.98
Rate worst pain (d5)[a]	0.97	0.98	0.98	0.95	0.97
How much pain relief due to pain treatments and/or medications (d6)	0.96	0.95	0.99	0.97	0.97

NOTE: Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.
[a] Pearson correlation for rating of worst pain, which is on a 0–10 scale.

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) evaluate the generalizability of these performance results (see Tables A.52–A.55 in the appendix). No significant differences were found for interrater reliability of the Pain Interview data elements in these sensitivity analyses.

Table 6.6 shows percent agreement for the two versions combined overall and by setting. Overall, percent agreement was excellent, ranging from 96 to 99 percent. At the setting level, percent agreement was also excellent and ranged from 95 to 99 percent in HHAs, 96 to 100 percent in IRFs, 96 to 100 percent in LTCHs, and 95 to 100 percent in SNFs.

Table 6.6. Interrater Reliability—Percent Agreement for Pain Interview Data Elements

Data Element	HHA (*n* = 197)	IRF (*n* = 256)	LTCH (*n* = 232)	SNF (*n* = 268)	Overall (*n* = 953)
Any pain (d1)	99	99	100	100	99
How often experienced pain (d2)	98	99	97	98	98
How often pain made it hard to sleep (d3)	95	98	98	100	98
Offered rehab therapies (d4a)	99	100	100	98	99
How often limited rehab due to pain (d4b)	97	98	98	99	98
How often limited daily activities due to pain (d4c)	97	98	99	99	98
Rate worst pain (d5)	96	96	96	95	96
How much pain relief due to pain treatments and/or medications (d6)	98	97	99	98	98

Day 3, 5, and 7 Repeat Assessment Evaluation

Tables 6.7a and 6.7b summarize patterns of change across the repeat assessment days. Patterns in Table 6.7a are characterized as "no change" (scores stay the same across assessment days), "worsen" (steady decline over assessment days), "improve" (steady improvement across assessment days), and "fluctuate" (scores go up and down across assessment days). As described in more detail in Volume 3, repeat assessment data were collected on a subset of the National Beta Test admission sample of patients/residents according to setting-level target totals. For example, each participating HHA was asked to contribute five patients for repeat assessment. Inclusion in repeat assessment data collection depended on assessor ability to complete the initial assessment on Day 3, availability of the assessor to return for repeat assessments on Days 5 and 7, and willingness of the patient/resident to complete multiple assessments. For the Pain data elements, data were collected for 476 patients/residents on Days 3, 5, and 7: 95 in HHAs, 122 in IRFs, 69 in LTCHs, and 190 in SNFs. Overall, pain presence, frequency, interference with sleep, therapy and daily activities, and rating of worst pain were significantly associated with assessment day, such that patients/residents generally demonstrated improvement, or less pain, over time, although effect sizes for differences among days tended to be very small.

There was a significant association between assessment day and pain presence ($F_{(2,589)} = 13.79$, $p < 0.01$), but the Cohen's d values for mean differences between days were not greater than 0.2. Frequency of pain was also significantly associated with assessment day ($F_{(2,590)} = 22.53$, $p < 0.01$), such that, compared with Day 3, pain frequency was slightly lower on Day 7 (Cohen's $d = 0.24$). Pain interference with sleep was significantly associated with assessment day, $F_{(2,590)} = 24.67$, $p < 0.01$, such that, compared with Day 3, pain interference with sleep was lower on Day 5 (Cohen's $d = 0.26$) and Day 7 (Cohen's $d = 0.25$). Pain interference with therapy activities was significantly associated with assessment day ($F_{(2,560)} = 6.16$, $p < 0.01$), but the Cohen's d values for mean differences between days were not greater than 0.2. Pain interference with daily activities was significantly associated with assessment day ($F_{(2,588)} = 29.19$, $p < 0.01$), such that, compared with Day 3, pain interference with daily activities was lower on Day 5 (Cohen's $d = 0.26$) and Day 7 (Cohen's $d = 0.31$). Finally, the rating of worst pain (shown in Table 6.7b) was significantly associated with assessment day ($F_{(2,588)} = 29.19$, $p < 0.01$), such that, compared with Day 3, pain interference with daily activities was lower on Day 5 (Cohen's $d = 0.22$) and Day 7 (Cohen's $d = 0.33$). There were no significant differences with effect sizes greater than 0.2 for differences between days 5 and 7.

These results convey a consistent pattern of a significant *but small* effect of time on pain regardless of which version is used; pain and most measures of interference tend to taper off in the first week of the PAC stay.

Table 6.7a. Day 3, 5, and 7 Repeat Assessment Results for Pain Data Elements (percent)

Data Element	HHA (*n* = 95)	IRF (*n* = 122)	LTCH (*n* = 69)	SNF (*n* = 190)	Overall (*n* = 476)
Any pain (d1)					
No change	89	89	75	78	83
Worsen	3	1	3	3	2
Improve	6	8	19	12	11
Fluctuate	1	2	3	7	4
How often experienced pain (d2)[a]					
No change	64	75	75	64	68
Worsen	6	3	6	6	6
Improve	16	12	12	22	17
Fluctuate	14	9	7	7	9
How often pain made it hard to sleep (d3)[a]					
No change	72	72	62	70	70
Worsen	5	5	6	6	6
Improve	14	17	19	19	18
Fluctuate	9	6	13	4	7
How often limited rehab due to pain (d4b)[a]					
No change	94	90	76	88	88
Worsen	0	1	8	6	4
Improve	6	6	10	2	5
Fluctuate	0	3	6	4	3
How often limited daily activities due to pain (d4c)[a]					
No change	77	71	69	64	70
Worsen	5	2	5	5	4
Improve	6	16	14	21	16
Fluctuate	11	11	12	9	10
How much pain relief due to pain treatments and/or medications (d6)					
No change	59	66	69	60	63
Improve	17	14	12	11	13
Worsen	14	14	7	15	14
Fluctuate	10	6	12	14	11

[a] Dichotomized for repeat assessments. For d2, d3, d4b, and d4c, we are looking at changes between "Rarely or not at all" and "Occasionally" versus "Frequently" and "Almost constantly." For d6, we are looking at changes between "No relief" and "Some relief" versus "Quite a bit of relief" and "Very much relief."

Table 6.7b. Day 3, 5, and 7 Repeat Assessment Results—Rating for Worst Pain

Data Element	HHA (*n* = 95)	IRF (*n* = 122)	LTCH (*n* = 69)	SNF (*n* = 190)	Overall (*n* = 476)
Rate worst pain (d5)					
Day 3 mean (SD)	6.39 (3.09)	5.99 (3.84)	7.01 (3.38)	6.36 (3.51)	6.37 (3.51)
Day 5 mean (SD)	5.64 (3.03)	5.29 (3.72)	6.65 (3.44)	5.20 (3.57)	5.52 (3.51)
Day 7 mean (SD)	5.12 (3.13)	5.10 (3.76)	5.87 (4.03)	4.96 (3.53)	5.16 (3.60)

Admission to Discharge

Tables 6.8a and 6.8b summarize patterns of change on Pain data elements from admission to discharge. As described in more detail in Volume 3, discharge data were collected on a subset of the National Beta Test admission sample of patients/residents. Availability of discharge data depended on advance notification of discharge and the ability to schedule assessments among the facility staff assessors in each participating site. Patterns are characterized as "no change" (scores stay the same at admission and discharge), "worsen" (scores decline from admission and discharge), and "improve" (scores improve from admission and discharge). For the Pain data elements, both admission and discharge data were collected on 784 patients/residents: 146 in HHAs, 333 in IRFs, 85 in LTCHs, and 220 in SNFs. Overall, responses to pain data elements tended to reflect improvement in pain presence, frequency, interference with sleep, daily activities, rehab, worst pain, and pain relief from admission to discharge. Specifically, compared with admission, at discharge patients/residents reported lower pain presence ($t_{(783)} = 6.95$, $p < 0.01$) less pain frequency ($t_{(460)} = 9.18$, $p < 0.01$), less interference with sleep ($t_{(459)} = 9.65$, $p < 0.01$), less interference with daily activities ($t_{(446)} = 8.98$, $p < 0.01$), less interference with rehab ($t_{(389)} = 3.99$, $p < 0.01$), lower worst pain ($t_{(453)} = 11.09$, $p < 0.01$), and an increase in pain relief due to treatment and/or medications ($t_{(418)} = 4.89$, $p < 0.01$).

Table 6.8a. Admission to Discharge Results for Pain Data Elements (percent)

Data Element	HHA (*n* = 146)	IRF (*n* = 333)	LTCH (*n* = 85)	SNF (*n* = 220)	Overall (*n* = 784)
Any pain (d1)					
No change	78	77	73	80	77
Worsen	4	6	11	3	5
Improve	18	17	16	17	17
How often experienced pain (d2)[a]					
No change	44	47	35	49	46
Worsen	15	11	17	13	13
Improve	40	43	48	38	41
How often pain made it hard to sleep (d3)[a]					

Data Element	HHA (*n* = 146)	IRF (*n* = 333)	LTCH (*n* = 85)	SNF (*n* = 220)	Overall (*n* = 784)
No change	53	45	29	47	45
Worsen	12	11	16	12	12
Improve	35	44	55	42	43
How often limited rehab due to pain (d4b)[a]					
No change	65	77	54	73	72
Worsen	9	8	13	7	8
Improve	26	15	33	19	19
How often limited daily activities due to pain (d4c)[a]					
No change	43	52	57	42	48
Worsen	8	12	17	13	12
Improve	48	37	26	46	40
How much pain relief due to pain treatments and/or medications (d6)					
No change	48	46	49	50	48
Worsen	18	16	20	19	18
Improve	33	38	31	31	35

[a] Dichotomized for repeat assessments. For d2, d3, d4b, and d4c, we are looking at changes between "Rarely or not at all" and "Occasionally" versus "Frequently" and "Almost constantly." For d6, we are looking at changes between "No relief" and "Some relief" versus "Quite a bit of relief" and "Very much relief."

Table 6.8b. Admission to Discharge Results—Rating for Worse Pain

	HHA (*n* = 82)	IRF (*n* = 207)	LTCH (*n* = 47)	SNF (*n* = 118)	Overall (*n* = 454)
Rate worst pain (d5)					
Admission Mean (SD)	6.37 (2.45)	7.71 (2.36)	7.85 (1.99)	7.66 (2.35)	7.47 (2.39)
Discharge Mean (SD)	5.28 (2.51)	6.34 (2.40)	6.79 (2.59)	6.27 (2.27)	6.18 (2.44)

Assessor Feedback

Feedback was strongly positive for the Pain Interview from facility staff in the focus groups and assessor survey. On the assessor survey, facility staff considered the Pain Interview to be the most clinically useful among all data elements, and research nurses rated it as the second-most clinically useful. In focus groups, both types of assessors agreed that the Pain data elements had high clinical utility, particularly for patient/resident transfer and care planning. Facility staff perceived this to be true because of their emphasis on improving or maintaining patient/resident functionality. They noted that pain affects mobility, sleep, and mood and each in turn "impacts their overall health." In fact, some facility staff went so far as to say that asking a patient for pain score alone without determining the impact of that pain is useless for clinical practice. Moreover,

the meaning and clinical utility of pain scores were critically dependent on understanding the patient's personal goals for pain management and daily functionality. Therefore, data elements about how pain affects sleep, therapy, and daily activities were well liked by facility staff in the focus groups. Although a small portion of facility staff and research nurses reported that the pain questions were redundant and confusing to some patients on the assessor survey, the majority considered the Pain data elements to carry low burden to assessors and patients/residents.

In the focus groups, both research nurses and facility staff noted that a weakness of the Pain data elements was that patients/residents experienced difficulty with recalling and characterizing pain experienced in the past.

> "How much of the time have you experienced pain or hurting over the last five days?" And they're going five days? I can't tell you what I did yesterday, and you ask me [about] five days ago?"
>
> —Durham, SNF staff

Data element reference time frame and patient/resident recall aside, these data elements were recognized as having high clinical relevance across PAC settings in both the survey and focus groups. In particular, questions about pain's impact on function were highly regarded.

Summary

Results for the Pain Interview data element set indicate very high overall support for standardization across PAC settings. Assessors reported that the Pain Interview had high clinical utility, particularly because of the emphasis on improving or maintaining patient/resident functionality. Results of the Day 3, 5, 7 repeat assessment indicate that pain generally improved across days, thus assessments conducted three days after admission may tend to identify slightly more pain than assessments conducted on Day 5 or later. Pain also generally improved from admission to discharge. Overall kappas were excellent, and percent agreement was high for both versions (past 3 days and past 5 days), with minimal setting differences. These combined results show high feasibility, excellent interrater reliability, reasonably low burden, and high clinical utility for the Pain Interview as a candidate data element for standardization across PAC settings.

7. Conclusion

The National Beta Test evaluated several candidate standardized patient assessment data elements in the clinical categories of Mental Status and Pain for possible use in the PAC assessment instruments. These data element sets include PHQ-2 to 9, PROMIS Depression, PROMIS Anxiety, and Pain Interview.

The general performance of these four data element sets is summarized for the combined sample in Table 7.1. As can be seen in Table 7.1, the four data element sets presented in this volume performed fairly well, showing feasibility, acceptable reliability, and moderate support from assessors. In terms of feasibility, missing data were very low for all four tested data element sets, and there was minimal variability in time to complete. PROMIS Depression (M = 2.2 minutes, SD = 0.8), PROMIS Anxiety (M = 2.2 minutes, SD = 0.8), and PHQ-2 to 9 data element sets (M = 2.3 minutes, SD = 1.5) were completed slightly more quickly than the Pain Interview (M = 2.6, SD = 1.4) data element set.

Interrater reliability was excellent for all four data element sets. These data elements also showed excellent interrater reliability based on percent agreement. The very strong reliability results are not surprising for interview-based assessment, as the interrater reliability was based on paired assessments that were conducted during a single patient interview. That is, both assessors were with the patient resident, one assessor conducted the interview, and both assessors recorded the patient's/resident's answer.

Stability or change from admission to discharge results (not shown in table) generally showed that there was some degree of stability, in that the majority of patients/residents did not exhibit significant change during their stay, but among those patients/residents who did change, symptoms of depression, anxiety, and pain improved over the course of the PAC stay. This improvement in symptoms at discharge implies that assessment of symptoms of depression, anxiety, and pain may be most informative at both admission and discharge to obtain a complete picture of a patient's/resident's mental status and pain during his or her PAC stay. We saw a similar trend of improvement for the Pain Interview data elements in the repeated assessments conducted on the same patient/resident on Days 3, 5, and 7. This supports the need for frequent, even daily, assessment of pain, which is commonly done in clinical practice.

Table 7.1. Summary of Mental Status and Pain Data Element Performance in National Beta Test (Combined Sample)

Data Element	Time to Complete (Mean, SD)	Interrater Reliability (Kappa)	Interrater Reliability (Percent Agreement)	Assessor Feedback
PHQ-2 to 9	2.3 (1.5)	0.95–1.00	97–100%	Moderate to high clinical utility, high data collection and patient burden
PROMIS Depression[a]	2.2 (0.8)	0.97–1.00	98–99%	Moderate clinical utility, relatively high data collection and patient burden
PROMIS Anxiety[a]	2.2 (0.8)	0.97–0.99	98–99%	Moderate clinical utility, relatively high data collection and patient burden
Pain Interview[b]	2.6 (1.4)	0.96–0.98	96–99%	High clinical utility, low burden

[a] Results for combined versions (past 3 days and past 7 days).
[b] Results for combined versions (past 3 days and past 5 days).
NOTE: Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.

Assessor feedback was somewhat mixed with regard to the Mental Status data elements and generally supportive of the Pain Interview data elements. All of the data elements were deemed at least moderately clinical useful by the clinical assessors in this study. Feedback from the clinical assessors in the National Beta Test indicated that the Pain Interview data elements had reasonably low burden, whereas the Mental Status data elements had relatively high data collection and patient burden. The assessors raised some other concerns with the Mental Status data elements that may be useful to consider in the future. Assessors noted that the personal nature of some of the data elements in the PHQ-2 to 9 will require training and may be uncomfortable for even well-trained staff, and that the questions in the PROMIS Depression data elements were upsetting to the patient/resident and that patients/residents may not be honest about their answers. The assessors also noted that the PROMIS Anxiety data elements were more palatable to patients/residents than the PHQ-2 to 9 and PROMIS Depression. These interview-based data elements were seen as somewhat more burdensome to patient/residents, perhaps because of the sensitive nature of the content. However, it is important to get the patient perspective, especially when assessing such clinical categories as mood and pain, as these are most reliably captured through direct patient/resident report.

Appendix. Supplementary Tables

Supplementary Tables for PHQ-2 to 9

Table A.1. Frequencies for PHQ-2 Positive Screen and Mean (SD) PHQ-9 Total Score by Patient/Resident Characteristics and Clinical Groups in the HHA Setting (percent)

Patient/Resident Characteristics and Clinical Groups	PHQ-2 Screen (Yes)	PHQ-9 Total Score M (SD)
Gender (n_{PHQ2} = 620[a]; n_{PHQ9} = 147)		
Male (n_{PHQ2} = 220; n_{PHQ9} = 40)	18.2	11.0 (5.2)
Female (n_{PHQ2} = 400; n_{PHQ9} = 107)	26.8	11.5 (5.1)
Age (n_{PHQ2} = 616[a]; n_{PHQ9} = 146)		
18–44 (n_{PHQ2} = 4; n_{PHQ9} = 1)	25.0	8.0 (—)
45–64 (n_{PHQ2} = 60; n_{PHQ9} = 24)	40.0	13.4 (4.9)
65–74 (n_{PHQ2} = 172; n_{PHQ9} = 49)	28.5	11.7 (5.2)
75–89 (n_{PHQ2} = 311; n_{PHQ9} = 59)	19.0	10.5 (4.7)
90+ (n_{PHQ2} = 69; n_{PHQ9} = 13)	18.8	10.7 (6.3)
Length of stay (n_{PHQ2} = 495; n_{PHQ9} = 119; mean, SD)	Pos: 31.0 (15.4) Neg: 31.2 (15.7)	Pearson r = 0.08
Disposition at discharge (n_{PHQ2} = 612; n_{PHQ9} = 145)		
Home (n_{PHQ2} = 455; n_{PHQ9} = 104)	22.9	10.7 (4.8)
Hospital (n_{PHQ2} = 23; n_{PHQ9} = 9)	39.1	14.5 (7.5)
Hospice (n_{PHQ2} = 12; n_{PHQ9} = 5)	41.7	13.6 (5.7)
HHA (n_{PHQ2} = 15; n_{PHQ9} = 3)	20.0	11.0 (1.0)
IRF (n_{PHQ2} = 4; n_{PHQ9} = 1)	25.0	4.0 (—)
LTCH (n_{PHQ2} = 1; n_{PHQ9} = 0)	0.0	—
SNF (n_{PHQ2} = 6; n_{PHQ9} = 1)	16.7	12.0 (—)
Other (n_{PHQ2} = 96; n_{PHQ9} = 22)	22.9	12.7 (5.2)
Clinical conditions (n_{PHQ2} = 417; n_{PHQ9} = 94)		
Sepsis		
Yes (n_{PHQ2} = 8; n_{PHQ9} = 0)	0.0	—
No (n_{PHQ2} = 409; n_{PHQ9} = 94)	23.0	11.0 (4.8)
Heart failure		
Yes (n_{PHQ2} = 32; n_{PHQ9} = 8)	25.0	11.6 (4.8)
No (n_{PHQ2} = 385; n_{PHQ9} = 86)	22.3	11.0 (4.8)
Stroke		
Yes (n_{PHQ2} = 6; n_{PHQ9} = 1)	16.7	10.0 (—)

Patient/Resident Characteristics and Clinical Groups	PHQ-2 Screen (Yes)	PHQ-9 Total Score M (SD)
No (n_{PHQ2} = 411; n_{PHQ9} = 93)	22.6	11.0 (4.8)
Mobility—Lying to sitting (n_{PHQ2} = 390; n_{PHQ9} = 90)		
Independent (n_{PHQ2} = 29; n_{PHQ9} = 6)	20.7	13.0 (3.8)
Setup or clean-up assistance (n_{PHQ2} = 59; n_{PHQ9} = 9)	15.3	10.0 (4.7)
Supervision or touching assistance (n_{PHQ2} = 119; n_{PHQ9} = 29)	24.4	11.5 (3.6)
Partial/moderate assistance (n_{PHQ2} = 125; n_{PHQ9} = 28)	22.4	11.4 (5.5)
Substantial/maximal assistance (n_{PHQ2} = 53; n_{PHQ9} = 16)	30.2	9.7 (5.9)
Dependent (n_{PHQ2} = 5; n_{PHQ9} = 2)	40.0	9.5 (7.8)

NOTE: Because of differences in sample sizes for PHQ-2 screening and PHQ-9 total scores, we report sample sizes for each (n_{PHQ2} = PHQ-2 screening; n_{PHQ9} = PHQ-9 total score).
[a] Significant ($p < 0.05$) associations with PHQ-2 positive screen as indicated by chi-square tests of independence.

Table A.2. Frequencies for PHQ-2 Positive Screen and Mean (SD) for PHQ-9 Total Score by Patient/Resident Characteristics and Clinical Groups in the IRF Setting (percent)

Patient/Resident Characteristics and Clinical Groups	PHQ-2 Positive Screen (Yes)	PHQ-9 Total Score M (SD)
Gender (n_{PHQ2} = 742; n_{PHQ9} = 202)		
Male (n_{PHQ2} = 318; n_{PHQ9} = 87)	27.4	11.1 (5.4)
Female (n_{PHQ2} = 424; n_{PHQ9} = 115)	27.1	12.4 (5.3)
Age (n_{PHQ2} = 739; n_{PHQ9} = 202)		
18–44 (n_{PHQ2} = 6; n_{PHQ9} = 3)	50.0	15.0 (3.6)
45–64 (n_{PHQ2} = 58; n_{PHQ9} = 22)	37.9	13.6 (6.2)
65–74 (n_{PHQ2} = 288; n_{PHQ9} = 79)	27.4	12.2 (5.9)
75–89 (n_{PHQ2} = 333; n_{PHQ9} = 89)	26.7	11.2 (4.6)
90+ (n_{PHQ2} = 54; n_{PHQ9} = 9)	16.7	9.4 (4.5)
Length of stay (n_{PHQ2} = 724; n_{PHQ9} = 197; mean, SD)	Pos: 14.2 (5.5) Neg: 14.1 (4.9)	Pearson r = –0.11
Disposition at discharge (n_{PHQ2} = 737[a]; n_{PHQ9} = 201)		
Home (n_{PHQ2} = 317; n_{PHQ9} = 72)	22.7	12.2 (5.1)
Hospital (n_{PHQ2} = 36; n_{PHQ9} = 15)	41.7	13.4 (4.9)
Hospice (n_{PHQ2} = 7; n_{PHQ9} = 0)	0.0	—
HHA (n_{PHQ2} = 252; n_{PHQ9} = 78)	31.0	11.8 (5.9)
IRF (n_{PHQ2} = 1; n_{PHQ9} = 0)	0.0	—
LTCH (n_{PHQ2} = 0; n_{PHQ9} = 0)	—	—
SNF (n_{PHQ2} = 104; n_{PHQ9} = 28)	26.9	11.1 (5.1)
Other (n_{PHQ2} = 20; n_{PHQ9} = 8)	40.0	8.0 (5.1)
Clinical conditions (n_{PHQ2} = *580;* n_{PHQ9} = 160)		
Sepsis		

Patient/Resident Characteristics and Clinical Groups	PHQ-2 Positive Screen (Yes)	PHQ-9 Total Score M (SD)
Yes (n_{PHQ2} = 25; n_{PHQ9} = 11)	44.0	10.8 (2.8)
No (n_{PHQ2} = 555; n_{PHQ9} = 149)	26.9	11.6 (5.1)
Heart failure		
Yes (n_{PHQ2} = 130[a]; n_{PHQ9} = 45)	34.6	12.2 (5.1)
No (n_{PHQ2} = 450; n_{PHQ9} = 115)	25.6	11.3 (4.9)
Stroke		
Yes (n_{PHQ2} = 98[a]; n_{PHQ9} = 18)	18.4	10.4 (3.8)
No (n_{PHQ2} = 482; n_{PHQ9} = 142)	29.5	11.7 (5.1)
Hygiene—Toileting (n_{PHQ2} = 564; n_{PHQ9} = 155)		
Independent (n_{PHQ2} = 5; n_{PHQ9} = 1)	20.0	6.0 (—)
Setup or clean-up assistance (n_{PHQ2} = 21; n_{PHQ9} = 1)	4.8	9.0 (—)
Supervision or touching assistance (n_{PHQ2} = 122; n_{PHQ9} = 30)	24.6	11.2 (4.3)
Partial/moderate assistance (n_{PHQ2} = 134; n_{PHQ9} = 40)	29.9	10.4 (5.3)
Substantial/maximal assistance (n_{PHQ2} = 139; n_{PHQ9} = 41)	29.5	12.3 (4.7)
Dependent (n_{PHQ2} = 143; n_{PHQ9} = 42)	29.4	12.1 (5.0)
Mobility—Lying to sitting (n_{PHQ2} = 577; n_{PHQ9} = 159)		
Independent (nPHQ2 = 42; nPHQ9 = 9)	21.4	10.8 (4.5)
Setup or clean-up assistance (nPHQ2 = 16; nPHQ9 = 6)	37.5	13.0 (4.6)
Supervision or touching assistance (nPHQ2 = 185; nPHQ9 = 55)	29.7	10.5 (4.2)
Partial/moderate assistance (nPHQ2 = 216; nPHQ9 = 52)	24.1	11.7 (5.2)
Substantial/maximal assistance (nPHQ2 = 91; nPHQ9 = 26)	28.6	12.9 (6.1)
Dependent (n_{PHQ2} = 27; n_{PHQ9} = 11)	40.7	11.8 (5.4)

NOTE: Because of differences in sample sizes for PHQ-2 screening and PHQ-9 total scores, we report sample sizes for each (n_{PHQ2} = PHQ-2 screening; n_{PHQ9} = PHQ-9 total score).
[a] Significant ($p < 0.05$) associations with PHQ-2 positive screen as indicated by chi-square tests of independence.

Table A.3. Frequencies for PHQ-2 Positive Screen and Mean (SD) PHQ-9 Total Score by Patient/Resident Characteristics and Clinical Groups in the LTCH Setting (percent)

Patient/Resident Characteristics and Clinical Groups	PHQ-2 Positive Screen (Yes)	PHQ-9 Total Score M (SD)
Gender (n_{PHQ2} = 458; n_{PHQ9} = 171)		
Male (n_{PHQ2} = 238; n_{PHQ9} = 87)	36.6	13.4 (6.1)
Female (n_{PHQ2} = 220; n_{PHQ9} = 84)	38.2	12.7 (5.7)
Age (n_{PHQ2} = 459; n_{PHQ9} = 172)		
18–44 (n_{PHQ2} = 23; n_{PHQ9} = 7)	30.4	13.1 (7.6)
45–64 (n_{PHQ2} = 113; n_{PHQ9} = 46)	40.7	13.2 (5.9)
65–74 (n_{PHQ2} = 165; n_{PHQ9} = 60)	36.4	12.6 (6.2)

Patient/Resident Characteristics and Clinical Groups	PHQ-2 Positive Screen (Yes)	PHQ-9 Total Score M (SD)
75–89 (n_{PHQ2} = 144; n_{PHQ9} = 57)	39.6	13.2 (5.5)
90+ (n_{PHQ2} = 14; n_{PHQ9} = 2)	14.3	18.5 (3.5)
Length of stay (n_{PHQ2} = 406; n_{PHQ9} = 150; mean, SD)	Pos: 24.6 (11.5) Neg: 23.2 (10.7)	Pearson r = –0.03
Disposition at discharge (n_{PHQ2} = 443; n_{PHQ9} = 166)		
Home (n_{PHQ2} = 90; n_{PHQ9} = 28)	31.1	12.5 (5.9)
Hospital (n_{PHQ2} = 31; n_{PHQ9} = 9)	29.0	12.3 (5.7)
Hospice (n_{PHQ2} = 12; n_{PHQ9} = 4)	33.3	13.5 (4.8)
HHA (n_{PHQ2} = 77; n_{PHQ9} = 30)	39.0	14.3 (6.5)
IRF (n_{PHQ2} = 45; n_{PHQ9} = 19)	42.2	15.7 (5.8)
LTCH (n_{PHQ2} = 1; n_{PHQ9} = 1)	100.0	14.6 (—)
SNF (n_{PHQ2} = 126; n_{PHQ9} = 50)	39.7	11.9 (5.7)
Other (n_{PHQ2} = 61; n_{PHQ9} = 25)	41.0	12.9 (6.3)
Clinical conditions (n_{PHQ2} = 390; n_{PHQ9} = 147)		
Sepsis		
Yes (n_{PHQ2} = 67; n_{PHQ9} = 27)	40.3	14.9 (6.2)
No (n_{PHQ2} = 323; n_{PHQ9} = 120)	37.2	12.9 (5.9)
Heart failure		
Yes (n_{PHQ2} = 13; n_{PHQ9} = 2)	15.4	17.0 (12.7)
No (n_{PHQ2} = 377; n_{PHQ9} = 145)	38.5	13.2 (5.9)
Stroke		
Yes (n_{PHQ2} = 26; n_{PHQ9} = 9)	34.6	13.0 (7.3)
No (n_{PHQ2} = 364; n_{PHQ9} = 138)	37.9	13.3 (5.9)
Hygiene—Toileting (n_{PHQ2} = 377[a]; n_{PHQ9} = 145)		
Independent (n_{PHQ2} = 46; n_{PHQ9} = 13)	28.3	14.1 (6.0)
Setup or clean-up assistance (n_{PHQ2} = 33; n_{PHQ9} = 4)	12.1	13.5 (7.5)
Supervision or touching assistance (n_{PHQ2} = 57; n_{PHQ9} = 24)	42.1	12.4 (6.3)
Partial/moderate assistance (n_{PHQ2} = 127; n_{PHQ9} = 24)	47.1	12.6 (6.3)
Substantial/maximal assistance (n_{PHQ2} = 63; n_{PHQ9} = 27)	42.9	12.4 (5.6)
Dependent (n_{PHQ2} = 51; n_{PHQ9} = 53)	41.7	14.3 (5.9)
Mobility—Lying to sitting (n_{PHQ2} = 340[a]; n_{PHQ9} = 131)		
Independent (n_{PHQ2} = 63; n_{PHQ9} = 21)	33.3	13.7 (6.9)
Setup or clean-up assistance (n_{PHQ2} = 24; n_{PHQ9} = 6)	25.0	14.5 (7.6)
Supervision or touching assistance (n_{PHQ2} = 58; n_{PHQ9} = 14)	24.1	13.5 (7.1)
Partial/moderate assistance (n_{PHQ2} = 75; n_{PHQ9} = 36)	50.7	12.6 (5.2)
Substantial/maximal assistance (n_{PHQ2} = 49; n_{PHQ9} = 23)	46.9	13.5 (4.7)
Dependent (n_{PHQ2} = 75; n_{PHQ9} = 31)	41.3	13.1 (6.6)

NOTE: Because of differences in sample sizes for PHQ-2 screening and PHQ-9 total scores, we report sample sizes for each (n_{PHQ2} = PHQ-2 screening; n_{PHQ9} = PHQ-9 total score).
[a] Significant ($p < 0.05$) associations with PHQ-2 positive screen as indicated by chi-square tests of independence.

Table A.4. Frequencies for PHQ-2 Positive Screen and Mean (SD) PHQ-9 Total Score by Patient/Resident Characteristics and Clinical Groups in the SNF Setting (percent)

Patient/Resident Characteristics and Clinical Groups	PHQ-2 Positive Screen (Yes)	PHQ-9 Total Score M (SD)
Gender (n_{PHQ2} = 1,086[a]; n_{PHQ9} = 296)		
Male (n_{PHQ2} = 423; n_{PHQ9} = 98)	23.4	11.8 (5.3)
Female (n_{PHQ2} = 663; n_{PHQ9} = 198)	30.3	11.5 (5.1)
Age (n_{PHQ2} = 1,081[a]; n_{PHQ9} = 294[b])		
18–44 (n_{PHQ2} = 9; n_{PHQ9} = 5)	55.6	14.0 (4.6)
45–64 (n_{PHQ2} = 73; n_{PHQ9} = 30)	41.1	14.4 (6.4)
65–74 (n_{PHQ2} = 284; n_{PHQ9} = 83)	29.2	11.8 (4.7)
75–89 (n_{PHQ2} = 543; n_{PHQ9} = 146)	27.4	11.0 (5.0)
90+ (n_{PHQ2} = 172; n_{PHQ9} = 30)	18.0	10.9 (4.8)
Length of stay (n_{PHQ2} = 941; n_{PHQ9} = 255; mean, SD)	Pos: 21.2 (11.6) Neg: 21.4 (12.5)	Pearson r = 0.05
Disposition at discharge (n_{PHQ2} = 1,064; n_{PHQ9} = 288)		
Home (n_{PHQ2} = 475; n_{PHQ9} = 123)	26.1	12.0 (5.1)
Hospital (n_{PHQ2} = 110; n_{PHQ9} = 34)	32.7	12.8 (6.0)
Hospice (n_{PHQ2} = 9; n_{PHQ9} = 0)	0.0	—
HHA (n_{PHQ2} = 277; n_{PHQ9} = 81)	29.6	10.7 (5.1)
IRF (n_{PHQ2} = 1; n_{PHQ9} = 0)	0.0	—
LTCH (n_{PHQ2} = 10; n_{PHQ9} = 1)	10.0	15.0 (—)
SNF (n_{PHQ2} = 43; n_{PHQ9} = 12)	27.9	10.2 (4.6)
Other (n_{PHQ2} = 139; n_{PHQ9} = 37)	26.6	11.7 (4.8)
Clinical conditions (n_{PHQ2} = *852;* n_{PHQ9} = 234)		
Sepsis		
Yes (n_{PHQ2} = 51; n_{PHQ9} = 17)	33.3	12.5 (5.8)
No (n_{PHQ2} = 801; n_{PHQ9} = 217)	27.5	11.8 (5.2)
Heart failure		
Yes (n_{PHQ2} = 210; n_{PHQ9} = 51)	24.3	12.0 (4.8)
No (n_{PHQ2} = 642; n_{PHQ9} = 183)	29.0	11.8 (5.3)
Stroke		
Yes (n_{PHQ2} = 62; n_{PHQ9} = 19)	30.7	13.1 (6.1)
No (n_{PHQ2} = 790; n_{PHQ9} = 215)	27.6	11.7 (5.1)
Hygiene—Toileting (n_{PHQ2} = 569[a]; n_{PHQ9} = 153)		
Independent (n_{PHQ2} = 22; n_{PHQ9} = 8)	36.4	11.6 (6.4)

Patient/Resident Characteristics and Clinical Groups	PHQ-2 Positive Screen (Yes)	PHQ-9 Total Score M (SD)
Setup or clean-up assistance (n_{PHQ2} = 23; n_{PHQ9} = 6)	26.1	9.8 (1.9)
Supervision or touching assistance (n_{PHQ2} = 143; n_{PHQ9} = 46)	32.2	11.5 (4.7)
Partial/moderate assistance (n_{PHQ2} = 74; n_{PHQ9} = 38)	22.4	12.3 (4.9)
Substantial/maximal assistance (n_{PHQ2} = 133; n_{PHQ9} = 27)	21.1	12.5 (5.2)
Dependent (n_{PHQ2} = 174; n_{PHQ9} = 28)	37.8	12.0 (6.4)
Mobility—Lying to sitting (n_{PHQ2} = 561; n_{PHQ9} = 149)		
Independent (n_{PHQ2} = 58; n_{PHQ9} = 16)	27.6	11.3 (5.6)
Setup or clean-up assistance (n_{PHQ2} = 14; n_{PHQ9} = 4)	28.6	8.25 (3.30)
Supervision or touching assistance (n_{PHQ2} = 165; n_{PHQ9} = 42)	25.5	11.7 (4.8)
Partial/moderate assistance (n_{PHQ2} = 200; n_{PHQ9} = 53)	27.5	12.0 (4.7)
Substantial/maximal assistance (n_{PHQ2} = 98; n_{PHQ9} = 26)	26.5	11.5 (5.2)
Dependent (n_{PHQ2} = 26; n_{PHQ9} = 8)	30.8	16.1 (8.2)

NOTE: Because of differences in sample sizes for PHQ-2 screening and PHQ-9 total scores, we report sample sizes for each (n_{PHQ2} = PHQ-2 screening; n_{PHQ9} = PHQ-9 total score).
[a] Significant ($p < 0.05$) associations with PHQ-2 positive screen as indicated by chi-square tests of independence.

Table A.5. Time to Complete PHQ-2 to 9 by Urbanicity (minutes)

	Urban (n = 1,621)	Nonurban (n = 106)	Overall (n = 1,727)
All			
Mean (SD)	2.3 (1.5)	2.1 (1.5)	2.3 (1.5)
PHQ-2 only			
Mean (SD)	1.7 (1.1)	1.6 (1.0)	1.7 (1.1)
PHQ-2 to 9			
Mean (SD)	3.9 (1.2)	4.2 (1.2)	4.0 (1.2)

Table A.6. Time to Complete PHQ-2 to 9 by Region (minutes)

	Northeast (n = 464)	South (n = 633)	Midwest (n = 357)	West (n = 273)	Overall (n = 1,727)
All					
Mean (SD)	2.2 (1.4)	2.4 (1.6)	2.5 (1.6)	2.2 (1.4)	2.3 (1.5)
PHQ-2 only					
Mean (SD)	1.7 (1.0)	1.7 (1.1)	1.8 (1.3)	1.7 (1.0)	1.7 (1.1)
PHQ-2 to 9					
Mean (SD)	3.8 (1.2)	4.0 (1.2)	4.2 (1.3)	3.9 (1.2)	4.0 (1.2)

Table A.7. Time to Complete PHQ-2 to 9 Data Elements by Facility Ownership (minutes)

	For-Profit (*n* = 1,063)	Nonprofit (*n* = 647)	Overall (*n* = 1,727[a])
All			
Mean (SD)	2.3 (1.5)	2.2 (1.4)	2.3 (1.5)
PHQ-2 Only			
Mean (SD)	1.7 (1.1)	1.7 (1.1)	1.7 (1.1)
PHQ-2 to 9			
Mean (SD)	4.0 (1.2)	3.8 (1.3)	4.0 (1.2)

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

Table A.8. Time to Complete PHQ-2 to 9 Data Elements by Facility Size (minutes)

	Below Setting-Type Median (*n* = 1,000)	Above Setting-Type Median (*n* = 726)	Overall (*n* = 1,727[a])
All			
Mean (SD)	2.3 (1.6)	2.2 (1.4)	2.3 (1.5)
PHQ-2 Only			
Mean (SD)	1.7 (1.2)	1.7 (1.0)	1.7 (1.1)
PHQ-2 to 9			
Mean (SD)	4.0 (1.2)	3.9 (1.3)	4.0 (1.2)

[a] 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.9. Interrater Reliability Kappa or Weighted Kappa for PHQ-2 to 9 Data Elements by Urbanicity

Data Element	Urban (*n* = 882)	Nonurban (*n* = 66)
Symptom present: little interest or pleasure (e1a1)	0.98	1.00
Symptom frequency: little interest or pleasure (e1a2)	0.99	1.00
Symptom present: feeling down, depressed, hopeless (e1b1)	0.99	1.00
Symptom frequency: feeling down, depressed, hopeless (e1b2)	0.98	1.00
Eligible for PHQ-9 per PHQ-2	0.98	0.96
Symptom present: too little/too much sleep (e1c1)	0.98	1.00
Symptom frequency: too little/too much sleep (e1c2)	0.97	1.00
Symptom present: tired/no energy (e1d1)	0.95	1.00
Symptom frequency: tired/no energy (e1d2)	0.98	1.00
Symptom present: poor appetite or overeating (e1e1)	0.96	1.00
Symptom frequency: poor appetite or overeating (e1e2)	0.97	1.00
Symptom present: feel bad about self (e1f1)	1.00	1.00

Data Element	Urban (*n* = 882)	Nonurban (*n* = 66)
Symptom frequency: feel bad about self (e1f2)	0.99	1.00
Symptom present: trouble concentrating (e1g1)	0.99	1.00
Symptom frequency: trouble concentrating (e1g2)	0.98	1.00
Symptom present: moving or speaking slowly (e1h1)	0.95	1.00
Symptom frequency: moving or speaking slowly (e1h2)	0.96	1.00
Symptom present: suicidal thoughts (e1i1)	0.98	1.00
Symptom frequency: suicidal thoughts (e1i2)	0.98	1.00
Sum of all symptom frequencies (PHQ-9)[a]	0.96	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, 0.81–1.00 is excellent/almost perfect.
[a] As classified into the five categories shown in Table A.1.

Table A.10. Interrater Reliability Kappa or Weighted Kappa for PHQ-2 to 9 Data Elements by Region

Data Element	Northeast (*n* = 208)	South (*n* = 356)	Midwest (*n* = 207)	West (*n* = 177)
Symptom present: little interest or pleasure (e1a1)	1.00	0.98	0.99	0.95
Symptom frequency: little interest or pleasure (e1a2)	1.00	0.99	0.98	0.97
Symptom present: feeling down, depressed, hopeless (e1b1)	0.99	1.00	0.99	0.98
Symptom frequency: feeling down, depressed, hopeless (e1b2)	0.99	0.99	0.99	0.96
Eligible for PHQ-9 per PHQ-2	0.99	0.99	0.96	0.96
Symptom present: too little/too much sleep (e1c1)	1.00	0.98	1.00	0.95
Symptom frequency: too little/too much sleep (e1c2)	0.99	0.96	0.99	0.97
Symptom present: tired/no energy (e1d1)	0.91	0.96	1.00	0.90
Symptom frequency: tired/no energy (e1d2)	0.99	0.97	1.00	0.96
Symptom present: poor appetite or overeating (e1e1)	0.97	0.94	1.00	0.96
Symptom frequency: poor appetite or overeating (e1e2)	0.97	0.96	1.00	1.00
Symptom present: feel bad about self (e1f1)	1.00	1.00	1.00	1.00
Symptom frequency: feel bad about self (e1f2)	1.00	0.98	1.00	1.00
Symptom present: trouble concentrating (e1g1)	1.00	1.00	0.96	1.00
Symptom frequency: trouble concentrating (e1g2)	0.96	0.99	0.96	1.00
Symptom present: moving or speaking slowly (e1h1)	0.94	0.94	1.00	0.96
Symptom frequency: moving or speaking slowly (e1h2)	0.95	0.95	1.00	0.97
Symptom present: suicidal thoughts (e1i1)	1.00	0.95	1.00	1.00
Symptom frequency: suicidal thoughts (e1i2)	0.96	0.97	1.00	1.00
Sum of all symptom frequencies (PHQ-9)[a]	0.95	0.96	0.95	0.97

NOTE: Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.
[a] As classified into the five categories shown in Table A.1.

Table A.11. Interrater Reliability Kappa or Weighted Kappa for PHQ-2 to 9 Data Elements by Facility Ownership

Data Element	For-Profit (*n* = 598)	Nonprofit (*n* = 344)
Symptom present: little interest or pleasure (e1a1)	0.98	1.00
Symptom frequency: little interest or pleasure (e1a2)	0.98	0.99
Symptom present: feeling down, depressed, hopeless (e1b1)	1.00	0.98
Symptom frequency: feeling down, depressed, hopeless (e1b2)	0.99	0.98
Eligible for PHQ-9 per PHQ-2	0.98	0.97
Symptom present: too little/too much sleep (e1c1)	0.99	0.98
Symptom frequency: too little/too much sleep (e1c2)	0.97	0.98
Symptom present: tired/no energy (e1d1)	0.95	0.95
Symptom frequency: tired/no energy (e1d2)	0.98	0.97
Symptom present: poor appetite or overeating (e1e1)	0.95	0.98
Symptom frequency: poor appetite or overeating (e1e2)	0.97	0.98
Symptom present: feel bad about self (e1f1)	1.00	1.00
Symptom frequency: feel bad about self (e1f2)	0.99	1.00
Symptom present: trouble concentrating (e1g1)	1.00	0.98
Symptom frequency: trouble concentrating (e1g2)	0.99	0.97
Symptom present: moving or speaking slowly (e1h1)	0.94	0.98
Symptom frequency: moving or speaking slowly (e1h2)	0.97	0.95
Symptom present: suicidal thoughts (e1i1)	0.97	1.00
Symptom frequency: suicidal thoughts (e1i2)	0.97	0.99
Sum of all symptom frequencies (PHQ-9)[a]	0.96	0.95

NOTE: Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.
[a] As classified into the five categories shown in Table A.1.

Table A.12. Interrater Reliability Kappa or Weighted Kappa for PHQ-2 to 9 Data Elements by Facility Size

Data Element	Below Setting-Type Median (*n* = 421)	Above Setting-Type Median (*n* = 526)
Symptom present: little interest or pleasure (e1a1)	0.99	0.97
Symptom frequency: little interest or pleasure (e1a2)	1.00	0.98
Symptom present: feeling down, depressed, hopeless (e1b1)	1.00	0.99
Symptom frequency: feeling down, depressed, hopeless (e1b2)	0.99	0.98
Eligible for PHQ-9 per PHQ-2	0.98	0.98
Symptom present: too little/too much sleep (e1c1)	0.98	0.99
Symptom frequency: too little/too much sleep (e1c2)	0.98	0.97
Symptom present: tired/no energy (e1d1)	0.97	-

Data Element	Below Setting-Type Median (*n* = 421)	Above Setting-Type Median (*n* = 526)
Symptom frequency: tired/no energy (e1d2)	0.99	0.97
Symptom present: poor appetite or overeating (e1e1)	1.00	0.93
Symptom frequency: poor appetite or overeating (e1e2)	1.00	0.95
Symptom present: feel bad about self (e1f1)	1.00	1.00
Symptom frequency: feel bad about self (e1f2)	0.99	0.99
Symptom present: trouble concentrating (e1g1)	0.98	1.00
Symptom frequency: trouble concentrating (e1g2)	0.98	0.98
Symptom present: moving or speaking slowly (e1h1)	0.98	0.93
Symptom frequency: moving or speaking slowly (e1h2)	0.99	0.95
Symptom present: suicidal thoughts (e1i1)	1.00	0.96
Symptom frequency: suicidal thoughts (e1i2)	0.99	0.97
Sum of all symptom frequencies (PHQ-9)[a]	0.98	0.94

NOTE: Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.
[a] As classified into the five categories shown in Table A.1.

Supplementary Tables for PROMIS Depression

Table A.13. Comparison of PROMIS Depression Data Elements and Scale Scores

	Market		Statistics		
Market Group A (past 7 days); Market Group B (past 3 days)	**A**	**B**	**t-Value**	**p-Value**	**Cohen's *d***
I felt worthless			1.92	0.06	0.07
Number of assessments	1,479	1,475			
Mean score	1.69	1.77			
Standard deviation	1.06	1.10			
I felt that I had nothing to look forward to			0.71	0.48	0.03
Number of assessments	1,468	1,469			
Mean score	1.66	1.69			
Standard deviation	1.05	1.05			
I felt hopeless			–2.00	0.05	-0.07
Number of assessments	1,480	1,468			
Mean score	2.18	2.08			
Standard deviation	1.36	1.23			
I felt sad			–1.02	0.31	-0.04
Number of assessments	1,485	1,476			
Mean score	2.26	2.21			

Market Group A (past 7 days); Market Group B (past 3 days)	Market		Statistics		
	A	B	t-Value	p-Value	Cohen's *d*
Standard deviation	1.18	1.19			
I felt lonely			0.69	0.49	0.03
Number of assessments	1,485	1,475			
Mean score	1.99	2.02			
Standard deviation	1.21	1.21			
I felt depressed			0.63	0.53	0.02
Number of assessments	1,486	1,466			
Mean score	1.97	2.00			
Standard deviation	1.19	1.18			
I felt I had no reason to live			0.96	0.34	0.04
Number of assessments	1,466	1,468			
Mean score	1.33	1.36			
Standard deviation	0.82	0.85			
I felt hopeless			–0.47	0.64	-0.02
Number of assessments	1,478	1,472			
Mean score	1.57	1.55			
Standard deviation	1.01	1.00			
PROMIS Depression total score			0.17	0.87	0.01
Number of assessments	1,498	1,488			
Mean score	14.59	14.63			
Standard deviation	6.87	6.99			
Interquartile range	6.87	6.99			
PROMIS Depression T-score			0.12	0.91	0.00
Number of assessments	1,498	1,488			
Mean score	51.08	51.12			
Standard deviation	9.67	9.72			
Interquartile range	9.67	9.72			

Table A.14. DIF Evaluation Pseudo R^2 for PROMIS Depression Data Elements According to Version, Setting, Gender, and Age

In the past 3/7 days,	Version	Setting	Gender	Age
I felt worthless	0.0019	0.0007	0.0001	0.0001
I felt that I had nothing to look forward to	0.0029	0.0005	0.0003	0.0002
I felt helpless	0.0000	0.0051	0.0046	0.0002
I felt sad	0.0000	0.0006	0.0023	0.0017
I felt lonely	0.0001	0.0026	0.0005	0.0048
I felt depressed	0.0001	0.0002	0.0006	0.0050
I felt I had no reason for living	0.0001	0.0015	0.0009	0.0005
I felt hopeless	0.0008	0.0004	0.0002	0.0007

NOTE: Age is dichotomized as less than or equal to 65 vs greater than 65.

Table A.15. Mean (SD) PROMIS Depression T-Score by Patient/Resident Characteristics and Clinical Groups in the HHA Setting

Patient/Resident Characteristics and Clinical Groups	PROMIS Depression T-Score M (SD)
Gender (*n* = 618[a])	
Male (*n* = 221)	47.9 (8.9)
Female (*n* = 397)	50.4 (9.3)
Age (*n* = 614[a])	
18–44 (*n* = 4)	51.9 (8.0)
45–64 (*n* = 60)	53.2 (9.3)
65–74 (*n* = 172)	50.4 (9.9)
75–89 (*n* = 310)	48.3 (8.6)
90+ (*n* = 68)	49.8 (9.1)
Length of stay (*n* = 494)	Pearson *r* = 0.07
Disposition at discharge (*n* = 610)	
Home (*n* = 454)	49.1 (9.2)
Hospital (*n* = 23)	52.5 (9.9)
Hospice (*n* = 12)	49.5 (8.9)
HHA (*n* = 15)	49.2 (10.0)
IRF (*n* = 4)	43.0 (10.0)
LTCH (*n* = 1)	53.3 (—)
SNF (*n* = 6)	52.0 (11.5)
Other (*n* = 95)	50.5 (9.0)
Clinical conditions (*n* = 416)	
Sepsis	
Yes (*n* = 8[a])	43.2 (6.7)

Patient/Resident Characteristics and Clinical Groups	PROMIS Depression T-Score M (SD)
No (*n* = 408)	49.6 (9.0)
Heart failure	
Yes (*n* = 32)	49.1 (9.9)
No (*n* = 384)	49.5 (8.9)
Stroke	
Yes (*n* = 6)	46.9 (9.9)
No (*n* = 410)	49.5 (9.0)
Mobility—Lying to sitting (*n* = 389)	
Independent (*n* = 30)	48.2 (8.6)
Setup or clean-up assistance (*n* = 58)	48.9 (8.9)
Supervision or touching assistance (*n* = 119)	49.5 (9.0)
Partial/moderate assistance (*n* = 125)	48.3 (9.2)
Substantial/maximal assistance (*n* = 52)	52.3 (8.9)
Dependent (*n* = 5)	47.5 (9.0)

[a] Indicates significant ($p < 0.05$) association between patient/resident characteristic and PROMIS Depression T-score.

Table A.16. Mean (SD) PROMIS Depression T-Score by Patient/Resident Characteristics and Clinical Groups in the IRF Setting

Patient/Resident Characteristics and Clinical Groups	PROMIS Depression T-Score M (SD)
Gender (*n* = 739)	
Male (*n* = 319)	50.7 (9.2)
Female (*n* = 420)	51.7 (9.4)
Age (*n* = 736[a])	
18–44 (*n* = *5)*	50.3 (13.8)
45–64 (*n* = 58)	56.5 (10.6)
65–74 (*n* = 286)	51.7 (9.2)
75–89 (*n* = 333)	50.3 (9.0)
90+ (*n* = 54)	49.9 (8.8)
Length of stay (*n* = 720)	Pearson *r* = 0.03
Disposition at discharge (*n* = 734)	
Home (*n* = 313)	51.0 (9.2)
Hospital (*n* = 36)	53.7 (11.8)
Hospice (*n* = 7)	51.3 (9.2)
HHA (*n* = 253)	51.7 (9.5)
IRF (*n* = 1)	48.3 (—)
LTCH (*n* = 0)	—

Patient/Resident Characteristics and Clinical Groups	PROMIS Depression T-Score M (SD)
SNF (*n* = 104)	50.3 (8.7)
Other (*n* = 20)	51.6 (7.8)
Clinical conditions (*n* = 576)	
Sepsis	
Yes (*n* = 24)	51.3 (8.9)
No (*n* = 552)	51.0 (9.5)
Heart failure	
Yes (*n* = 128)	50.7 (10.2)
No (*n* = 448)	51.1 (9.2)
Stroke	
Yes (*n* = 98)	50.5 (9.3)
No (*n* = 478)	41.1 (9.5)
Hygiene—Toileting (*n* = 560[a])	
Independent (*n* = *5)*	44.4 (7.8)
Setup or clean-up assistance (*n* = 21)	46.3 (7.6)
Supervision or touching assistance (*n* = 120)	48.5 (8.3)
Partial/moderate assistance (*n* = 135)	51.4 (10.1)
Substantial/maximal assistance (*n* = 139)	52.7 (9.8)
Dependent (*n* = 140)	51.9 (9.0)
Mobility—Lying to sitting (*n* = 573[a])	
Independent (*n* = 42)	48.1 (8.8)
Setup or clean-up assistance (*n* = 16)	51.9 (10.0)
Supervision or touching assistance (*n* = 185)	49.9 (9.2)
Partial/moderate assistance (*n* = 213)	51.7 (9.4)
Substantial/maximal assistance (*n* = 91)	51.8 (10.0)
Dependent (*n* = 26)	54.9 (9.4)

[a] Indicates significant ($p < 0.05$) association between patient/resident characteristic and PROMIS Depression T-score.

Table A.17. Mean (SD) PROMIS Depression T-Score by Patient/Resident Characteristics and Clinical Groups in the LTCH Setting

Patient/Resident Characteristics and Clinical Groups	PROMIS Depression T-Score M (SD)
Gender (*n* = 453)	
Male (*n* = 236)	53.1 (10.3)
Female (*n* = 217)	54.4 (10.1)
Age (*n* = 454)	
18–44 (*n* = 22)	53.9 (10.7)

Patient/Resident Characteristics and Clinical Groups	PROMIS Depression T-Score M (SD)
45–64 (*n* = 113)	54.4 (10.4)
65–74 (*n* = 162)	53.6 (10.2)
75–89 (*n* = 143)	53.7 (10.2)
90+ (*n* = 14)	50.5 (9.4)
Length of stay (*n* = 401)	Pearson *r* = 0.06
Disposition at discharge (*n* = 438)	
Home (*n* = 88)	52.7 (9.4)
Hospital (*n* = 31)	52.5 (9.5)
Hospice (*n* = 12)	50.9 (9.2)
HHA (*n* = 77)	54.0 (10.1)
IRF (*n* = 44)	55.5 (11.0)
LTCH (*n* = 1)	38.0 (—)
SNF (*n* = 124)	54.2 (11.0)
Other (*n* = 61)	53.9 (9.4)
Clinical conditions (*n* = 387)	
Sepsis	
Yes (*n* = 66[a])	56.5 (11.3)
No (*n* = 321)	53.3 (9.8)
Heart failure	
Yes (*n* = 13)	51.2 (12.1)
No (*n* = 374)	53.9 (10.1)
Stroke	
Yes (*n* = 25)	54.3 (12.0)
No (*n* = 362)	53.8 (10.0)
Hygiene—Toileting (*n* = 374)	
Independent (*n* = 46)	52.7 (11.7)
Setup or clean-up assistance (*n* = 33)	49.7 (7.9)
Supervision or touching assistance (*n* = 57)	52.5 (8.9)
Partial/moderate assistance (*n* = 51)	54.8 (8.9)
Substantial/maximal assistance (*n* = 62)	54.9 (10.3)
Dependent (*n* = 125)	55.0 (11.0)
Mobility—Lying to sitting (*n* = 340[a])	
Independent (*n* = 63)	53.7 (11.1)
Setup or clean-up assistance (*n* = 24)	52.8 (9.7)
Supervision or touching assistance (*n* = 58)	50.5 (9.4)
Partial/moderate assistance (*n* = 71)	53.7 (8.6)
Substantial/maximal assistance (*n* = 48)	57.6 (10.5)

Patient/Resident Characteristics and Clinical Groups	PROMIS Depression T-Score M (SD)
Dependent (*n* = 76)	54.4 (10.4)

[a] Indicates significant ($p < 0.05$) association between patient/resident characteristic and PROMIS Depression T-score.

Table A.18. Frequencies for PROMIS Depression T-Score by Patient/Resident Characteristics and Clinical Groups in the SNF Setting (percent)

Patient/Resident Characteristics and Clinical Groups	PROMIS Depression T-Score M (SD)
Gender (*n* = 1,074)	
Male (*n* = 420)	50.2 (9.9)
Female (*n* = 654)	51.2 (9.7)
Age (*n* = 1,069[a])	
18–44 (*n* = 8)	57.4 (13.7)
45–64 (*n* = 73)	54.7 (11.0)
65–74 (*n* = 282)	51.1 (9.6)
75–89 (*n* = 536)	50.4 (9.6)
90+ (*n* = 170)	49.8 (9.4)
Length of stay (*n* = 933)	Pearson *r* = 0.01
Disposition at discharge (*n* = 1,052)	
Home (*n* = 473)	50.1 (9.4)
Hospital (*n* = 106)	51.1 (10.5)
Hospice (*n* = 9)	48.7 (7.8)
HHA (*n* = 274)	51.2 (9.3)
IRF (*n* = 1)	49.2 (—)
LTCH (*n* = 10)	49.2 (10.6)
SNF (*n* = 43)	51.8 (11.3)
Other (*n* = 136)	52.0 (10.9)
Clinical conditions (*n* = 839)	
Sepsis	
Yes (*n* = 50)	51.9 (10.1)
No (*n* = 791)	51.0 (9.9)
Heart failure	
Yes (*n* = 206)	51.3 (9.4)
No (*n* = 635)	50.9 (10.0)
Stroke	
Yes (*n* = 63)	51.6 (11.3)
No (*n* = 778	51.0 (9.8)

Patient/Resident Characteristics and Clinical Groups	PROMIS Depression T-Score M (SD)
Hygiene—Toileting (*n* = 562)	
Independent (*n* = 21)	48.7 (7.1)
Setup or clean-up assistance (*n* = 23)	48.0 (9.1)
Supervision or touching assistance (*n* = 142)	49.3 (9.3)
Partial/moderate assistance (*n* = 174)	50.7 (9.8)
Substantial/maximal assistance (*n* = 131)	51.6 (9.4)
Dependent (*n* = 71)	51.4 (11.0)
Mobility—Lying to sitting (*n* = 554)	
Independent (*n* = 57)	48.8 (8.8)
Setup or clean-up assistance (*n* = 14)	51.6 (7.1)
Supervision or touching assistance (*n* = 165)	49.6 (9.1)
Partial/moderate assistance (*n* = 197)	51.3 (10.1)
Substantial/maximal assistance (*n* = 98)	50.8 (10.3)
Dependent (*n* = 23)	51.7 (9.3)

[a] Indicates significant ($p < 0.05$) association between patient/resident characteristic and PROMIS Depression T-score.

Table A.19. Time to Complete PROMIS Depression by Urbanicity (minutes)

	Urban (*n* = 1,432)	Nonurban (*n* = 95)	Overall (*n* = 1,527)
Mean (SD)	2.2 (0.8)	2.3 (0.9)	2.2 (0.8)

Table A.20. Time to Complete PROMIS Depression by Region (minutes)

	Northeast (*n* = 423)	South (*n* = 565)	Midwest (*n* = 304)	West (*n* = 235)	Overall (*n* = 1,527)
Mean (SD)	2.2 (0.8)	2.2 (0.9)	2.2 (0.9)	2.2 (0.8)	2.2 (0.8)

Table A.21. Time to Complete PROMIS Depression Data Elements by Facility Ownership (minutes)

	For-Profit (*n* = 920)	Nonprofit (*n* = 595)	Overall (*n* = 1,527[a])
Mean (SD)	2.2 (0.9)	2.2 (0.8)	2.2 (0.8)

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

Table A.22. Time to Complete PROMIS Depression Data Elements by Facility Size (minutes)

	Below Setting-Type Median (*n* = 655)	Above Setting-Type Median (*n* = 871)	Overall (*n* = 1,527[a])
Mean (SD)	2.3 (0.9)	2.2 (0.8)	2.2 (0.8)

[a] 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.23. Interrater Reliability Kappa or Weighted Kappa for PROMIS Depression Items by Urbanicity

Data Element	Urban (*n* = 869)	Nonurban (*n* = 66)
How often felt worthless (e2a)	0.98	1.00
How often felt there was nothing to look forward to (e2b)	0.97	1.00
How often felt helpless (e2c)	0.98	0.97
How often felt sad (e2d)	0.98	1.00
How often felt lonely (e2e)	0.99	1.00
How often felt depressed (e2f)	0.98	1.00
How often felt there was no reason for living (e2g)	0.97	1.00
How often felt hopeless (e2h)	0.98	1.00
Depression total[a]	0.99	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, 0.81–1.00 is excellent/almost perfect.
[a] Pearson correlation for total scores.

Table A.24. Interrater Reliability Kappa or Weighted Kappa for PROMIS Depression items by Region

Data Element	Northeast (*n* = 204)	South (*n* = 354)	Midwest (*n* = 205)	West (*n* = 172)
How often felt worthless (e2a)	0.99	0.99	0.99	0.98
How often felt there was nothing to look forward to (e2b)	0.95	0.99	0.98	0.95
How often felt helpless (e2c)	0.99	0.98	0.98	0.98
How often felt sad (e2d)	0.97	0.99	0.99	0.98
How often felt lonely (e2e)	0.97	0.99	0.99	1.00
How often felt depressed (e2f)	0.99	0.98	0.99	0.99
How often felt there was no reason for living (e2g)	0.99	0.98	0.97	0.98
How often felt hopeless (e2h)	0.97	0.99	0.98	0.98
Depression total[a]	0.98	1.00	0.99	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, 0.81–1.00 is excellent/almost perfect.
[a] Pearson correlation for total scores.

Table A.25. Interrater Reliability Kappa or Weighted Kappa for PROMIS Depression Data Elements by Facility Ownership

Data Element	For-Profit (*n* = 589)	Nonprofit (*n* = 340)
How often felt worthless (e2a)	0.99	0.98
How often felt there was nothing to look forward to (e2b)	0.99	0.96
How often felt helpless (e2c)	0.99	0.99
How often felt sad (e2d)	0.99	0.97
How often felt lonely (e2e)	0.99	0.97
How often felt depressed (e2f)	0.99	0.98
How often felt there was no reason for living (e2g)	0.98	0.97
How often felt hopeless (e2h)	0.98	0.99
Depression total[a]	0.99	0.99

NOTE: Interpretation of kappa or weighted kappa is as follows: 0.00–0.20: slight/poor, 0.21–0.40: fair, 0.41–0.60: moderate; 0.61–0.80: substantial/good, 0.81–1.00: excellent/almost perfect.
[a] Pearson correlation for total scores.

Table A.26. Interrater Reliability Kappa or Weighted Kappa for PROMIS Depression Data Elements by Facility Size

Data Element	Below Setting-Type Median (*n* = 419)	Above Setting-Type Median (*n* = 515)
How often felt worthless (e2a)	0.99	0.98
How often felt there was nothing to look forward to (e2b)	0.99	0.96
How often felt helpless (e2c)	0.99	0.98
How often felt sad (e2d)	0.99	0.97
How often felt lonely (e2e)	0.99	0.98
How often felt depressed (e2f)	0.99	0.98
How often felt there was no reason for living (e2g)	0.97	0.98
How often felt hopeless (e2h)	0.99	0.98
Depression total[a]	0.99	0.99

NOTE: Interpretation of kappa or weighted kappa is as follows: 0.00–0.20: slight/poor, 0.21–0.40: fair, 0.41–0.60: moderate; 0.61–0.80: substantial/good, 0.81–1.00: excellent/almost perfect.
[a] Pearson correlation for total scores.

Table A.27. Summed Score to T-Score Conversion Table for Eight PROMIS Depression Data Elements Included in National Beta Test.

Raw Summed Score	T-Score	Standard Error of T-Score
8	38.0	5.7
9	44.3	3.6
10	47.1	3.0
11	49.1	2.6
12	50.7	2.4
13	52.1	2.2
14	53.2	2.0
15	54.3	2.0
16	55.3	1.9
17	56.3	1.9
18	57.2	1.9
19	58.1	1.9
20	59.0	1.9
21	59.9	1.9
22	60.8	1.9
23	61.7	1.9
24	62.6	1.9
25	63.5	1.9
26	64.4	1.9
27	65.3	1.9
28	66.2	1.9
29	67.2	1.9
30	68.1	1.9
31	69.1	1.9
32	70.1	1.9
33	71.1	1.9
34	72.1	1.9
35	73.2	1.9
36	74.4	2.0
37	75.6	2.1
38	77.1	2.3
39	78.9	2.6
40	81.7	3.2

NOTE: This table is valid only for summed scores based on the eight items administered in the National Beta Test.

Supplementary Tables for PROMIS Anxiety

Table A.28. Comparison of PROMIS Anxiety Data Elements and Scale Scores According to Version

	Market		Statistics		
Market Group A (past 7 days); Market Group B (past 3 days)	**A**	**B**	**t-Value**	**p-Value**	**Cohen's *d***
I felt worried			–3.05	0.00	–0.11
Number of assessments	1,478	1,472			
Mean score	2.38	2.24			
Standard deviation	1.21	1.17			
My worries overwhelmed me			–2.14	0.03	–0.08
Number of assessments	1,477	1,470			
Mean score	1.74	1.65			
Standard deviation	1.09	1.05			
I had trouble paying attention			–0.22	0.83	–0.01
Number of assessments	1,475	1,468			
Mean score	1.89	1.88			
Standard deviation	1.08	1.09			
I felt nervous			–0.60	0.55	–0.02
Number of assessments	1,483	1,473			
Mean score	2.07	2.04			
Standard deviation	1.13	1.16			
I had difficulty calming down			–0.09	0.93	0.00
Number of assessments	1,473	1,463			
Mean score	1.69	1.69			
Standard deviation	1.02	1.02			
I found it hard to focus on anything other than my anxiety			0.08	0.94	0.00
Number of assessments	1,469	1,456			
Mean score	1.64	1.65			
Standard deviation	1.01	1.03			
I felt like I needed help for my anxiety			0.47	0.64	0.02
Number of assessments	1,465	1,458			
Mean score	1.69	1.71			
Standard deviation	1.10	1.12			
I had sudden feelings of panic			–0.61	0.54	–0.02
Number of assessments	1,472	1,470			

Market Group A (past 7 days); Market Group B (past 3 days)	Market		Statistics		
	A	B	t-Value	p-Value	Cohen's *d*
Mean score	1.48	1.46			
Standard deviation	0.91	0.91			
Anxiety total score			–0.98	0.33	–0.04
Number of assessments	1,492	1,479			
Mean score	14.52	14.28			
Standard deviation	6.60	6.63			
Interquartile range	14.52	14.28			
Anxiety T-score			–1.64	0.10	–0.06
Number of assessments	1,492	1,479			
Mean score	50.77	50.16			
Standard deviation	9.98	1.18			
Interquartile range	14.31	14.28			

Table A.29. DIF Evaluation Pseudo R^2 for PROMIS Anxiety Data Elements According to Version, Setting, Gender, and Age

In the past 3/7 days,	Version	Setting	Gender	Age
I felt worried	0.0015	0.0006	0.0033	0.0001
My worries overwhelmed me	0.0009	0.0015	0.0019	0.0012
I had trouble paying attention	0.0012	0.0012	0.0001	0.0000
I felt nervous	0.0006	0.0008	0.0060	0.0003
I had difficulty calming down	0.0002	0.0011	0.0003	0.0000
I found it hard to focus on anything other than my anxiety	0.0002	0.0017	0.0001	0.0015
I felt like I needed help for my anxiety	0.0005	0.0014	0.0004	0.0019
I had sudden feelings of panic	0.0010	0.0020	0.0002	0.0005

NOTE: Age is dichotomized as 65 or younger versus older than 65.

Table A.30. Mean (SD) PROMIS Anxiety T-Score by Patient/Resident Characteristics and Clinical Groups in the HHA Setting

Patient/Resident Characteristics and Clinical Groups	PROMIS Anxiety T-Score M (SD)
Gender (*n* = 618[a])	
Male (*n* = 221)	47.4 (9.1)
Female (*n* = 397)	50.5 (9.7)
Age (*n* = 614[a])	
18–44 (*n* = 4)	53.0 (9.2)
45–64 (*n* = 60)	53.6 (10.2)

Patient/Resident Characteristics and Clinical Groups	PROMIS Anxiety T-Score M (SD)
65–74 (*n* = 172)	51.5 (9.9)
75–89 (*n* = 310)	47.4 (8.9)
90+ (*n* = 68)	49.2 (9.6)
Length of stay (*n* = 494)	Pearson *r* = 0.003
Disposition at discharge (*n* = 610)	
Home (*n* = 454)	48.9 (9.5)
Hospital (*n* = 23)	52.6 (10.6)
Hospice (*n* = 12)	48.1 (11.9)
HHA (*n* = 15)	50.3 (9.8)
IRF (*n* = 4)	41.4 (8.8)
LTCH (*n* = 1)	44.2 (—)
SNF (*n* = 6)	47.7 (6.8)
Other (*n* = 95)	51.2 (9.8)
Clinical conditions (*n* = 416)	
Sepsis	
Yes (*n* = 8[a])	42.5 (6.2)
No (*n* = 408)	49.5 (9.4)
Heart failure	
Yes (*n* = 32)	49.2 (9.8)
No (*n* = 384)	49.4 (9.4)
Stroke	
Yes (*n* = 6)	47.4 (9.1)
No (*n* = 410)	49.4 (9.4)
Mobility—Lying to sitting (*n* = 389)	
Independent (*n* = 30)	48.2 (9.4)
Setup or clean-up assistance (*n* = 58)	49.2 (9.4)
Supervision or touching assistance (*n* = 119)	49.5 (9.5)
Partial/moderate assistance (*n* = 125)	48.2 (9.3)
Substantial/maximal assistance (*n* = 52)	51.5 (10.0)
Dependent (*n* = 5)	45.4 (7.0)

[a] Indicates significant ($p < 0.05$) association between patient/resident characteristic and PROMIS Anxiety T-score.

Table A.31. Mean (SD) PROMIS Anxiety T-Score by Patient/Resident Characteristics and Clinical Groups in the IRF Setting

Patient/Resident Characteristics and Clinical Groups	PROMIS Anxiety T-Score M (SD)
Gender (*n* = 737)	
Male (*n* = 317)	50.4 (9.8)
Female (*n* = 420)	51.6 (10.1)
Age (*n* = 734[a])	
18–44 (*n* = 5)	52.0 (12.5)
45–64 (*n* = 58)	56.9 (12.4)
65–74 (*n* = 286)	51.6 (9.8)
75–89 (*n* = 331)	49.8 (9.4)
90+ (*n* = 54)	49.6 (9.2)
Length of stay (*n* = 719)	Pearson *r* = 0.03
Disposition at discharge (*n* = 732)	
Home (*n* = 311)	50.3 (9.9)
Hospital (*n* = 36)	54.1 (11.2)
Hospice (*n* = 7)	55.1 (7.9)
HHA (*n* = 252)	51.5 (10.1)
IRF (*n* = 1)	48.9 (—)
LTCH (*n* = 0)	—
SNF (*n* = 105)	51.1 (9.6)
Other (*n* = 20)	50.1 (8.4)
Clinical conditions (*n* = 574)	
Sepsis	
Yes (*n* = 24)	52.5 (8.1)
No (*n* = 550)	50.8 (9.9)
Heart failure	
Yes (*n* = 129)	51.0 (11.0)
No (*n* = 445)	50.9 (9.5)
Stroke	
Yes (*n* = 98)	51.0 (9.2)
No (*n* = 476)	50.9 (10.0)
Hygiene—Toileting (*n* = 558[a])	
Independent (*n* = 5)	42.9 (9.4)
Setup or clean-up assistance (*n* = 21)	46.7 (8.7)
Supervision or touching assistance (*n* = 120)	49.4 (8.6)
Partial/moderate assistance (*n* = 134)	50.9 (10.1)
Substantial/maximal assistance (*n* = 139)	51.5 (9.8)
Dependent (*n* = 139)	52.3 (10.0)

Patient/Resident Characteristics and Clinical Groups	PROMIS Anxiety T-Score M (SD)
Mobility—Lying to sitting (*n* = 571[a])	
Independent (*n* = 42)	48.1 (8.8)
Setup or clean-up assistance (*n* = 16)	53.2 (13.1)
Supervision or touching assistance (*n* = 185)	49.6 (9.3)
Partial/moderate assistance (*n* = 212)	51.7 (9.5)
Substantial/maximal assistance (*n* = 91)	51.6 (10.9)
Dependent (*n* = 25)	54.6 (10.2)

[a] Indicates significant ($p < 0.05$) association between patient/resident characteristic and PROMIS Anxiety T-score.

Table A.32. Mean (SD) PROMIS Anxiety T-Score by Patient/Resident Characteristics and Clinical Groups in the LTCH Setting

Patient/Resident Characteristics and Clinical Groups	PROMIS Anxiety T-Score M (SD)
Gender (*n* = 450)	
Male (*n* = 234)	52.7 (10.9)
Female (*n* = 216)	54.4 (10.6)
Age (*n* = 451)	
18–44 (*n* = 22)	50.3 (11.2)
45–64 (*n* = 112)	54.4 (11.1)
65–74 (*n* = 162)	54.4 (10.8)
75–89 (*n* = 141)	52.7 (10.5)
90+ (*n* = 14)	49.9 (10.4)
Length of stay (*n* = 397)	Pearson *r* = 0.03
Disposition at discharge (*n* = 435)	
Home (*n* = 88)	51.6 (10.6)
Hospital (*n* = 31)	52.4 (7.6)
Hospice (*n* = 12)	52.3 (11.0)
HHA (*n* = 77)	53.8 (10.8)
IRF (*n* = 45)	54.7 (11.3)
LTCH (*n* = 1)	44.9 (—)
SNF (*n* = 120)	54.1 (11.2)
Other (*n* = 61)	54.9 (10.7)
Clinical conditions (*n* = 383)	
Sepsis	
Yes (*n* = 65[a])	56.8 (11.5)
No (*n* = 318)	53.0 (10.3)
Heart failure	
Yes (*n* = 13)	49.4 (8.8)

Patient/Resident Characteristics and Clinical Groups	PROMIS Anxiety T-Score M (SD)
No (*n* = 370)	53.8 (10.6)
Stroke	
Yes (*n* = 24)	51.8 (12.7)
No (*n* = 359)	53.8 (10.5)
Hygiene—Toileting (*n* = 371)	
Independent (*n* = 45)	51.4 (12.9)
Setup or clean-up assistance (*n* = 33)	50.1 (8.6)
Supervision or touching assistance (*n* = 57)	53.0 10.0)
Partial/moderate assistance (*n* = 51)	54.7 (9.7)
Substantial/maximal assistance (*n* = 61)	54.7 (9.6)
Dependent (*n* = 124)	55.1 (11.3)
Mobility—Lying to sitting (*n* = 337[a])	
Independent (*n* = 61)	51.8 (12.9)
Setup or clean-up assistance (*n* = 24)	52.5 (9.4)
Supervision or touching assistance (*n* = 58)	50.9 (8.8)
Partial/moderate assistance (*n* = 71)	55.1 (9.5)
Substantial/maximal assistance (*n* = 48)	56.8 (10.8)
Dependent (*n* = 75)	54.8 (10.3)

[a] Indicates significant ($p < 0.05$) association between patient/resident characteristic and PROMIS Anxiety T-score.

Table A.33. Mean (SD) PROMIS Anxiety T-Score by Patient/Resident Characteristics and Clinical Groups in the SNF Setting

Patient/Resident Characteristics and Clinical Groups	PROMIS Anxiety T-Score M (SD)
Gender (*n* = 1,065)	
Male (*n* = 416)	48.8 (9.9)
Female (*n* = 649)	49.9 (9.8)
Age (*n* = 1,060[a])	
18–44 (*n* = 8)	56.0 (12.8)
45–64 (*n* = 73)	54.0 (11.4)
65–74 (*n* = 280)	49.9 (9.9)
75–89 (*n* = 533)	49.1 (9.7)
90+ (*n* = 166)	47.8 (8.7)
Length of stay (*n* = 929)	Pearson *r* = 0.004
Disposition at discharge (*n* = 1,044)	
Home (*n* = 471)	49.4 (9.6)
Hospital (*n* = 104)	49.5 (10.9)
Hospice (*n* = 9)	44.3 (6.9)

Patient/Resident Characteristics and Clinical Groups	PROMIS Anxiety T-Score M (SD)
HHA (*n* = 274)	49.1 (9.5)
IRF (*n* = 1)	43.1 (—)
LTCH (*n* = 9)	50.4 (11.4)
SNF (*n* = 42)	51.1 (10.7)
Other (*n* = 134)	50.2 (10.4)
Clinical conditions (*n* = 836)	
Sepsis	
Yes (*n* = 50)	52.1 (11.4)
No (*n* = 786)	49.5 (9.9)
Heart failure	
Yes (*n* = 205)	49.2 (10.0)
No (*n* = 631)	49.8 (10.0)
Stroke	
Yes (*n* = 62)	51.1 (10.5)
No (*n* = 774)	49.5 (10.0)
Hygiene—Toileting (*n* = 559)	
Independent (*n* = 21)	47.3 (9.6)
Setup or clean-up assistance (*n* = 23)	46.0 (6.9)
Supervision or touching assistance (*n* = 141)	49.1 (9.7)
Partial/moderate assistance (*n* = 172)	48.6 (9.4)
Substantial/maximal assistance (*n* = 131)	49.9 (10.0)
Dependent (*n* = 71)	49.7 (10.7)
Mobility—Lying to sitting (*n* = 551)	
Independent (*n* = 56)	48.2 (9.4)
Setup or clean-up assistance (*n* = 14)	49.8 (7.5)
Supervision or touching assistance (*n* = 164)	48.5 (9.4)
Partial/moderate assistance (*n* = 196)	49.2 (9.8)
Substantial/maximal assistance (*n* = 97)	49.8 (10.3)
Dependent (*n* = 24)	50.1 (9.8)

[a] Indicates significant ($p < 0.05$) association between patient/resident characteristic and PROMIS Anxiety T-score.

Table A.34. Time to Complete PROMIS Anxiety by Urbanicity (minutes)

	Urban (*n* = 1,463)	Nonurban (*n* = 99)	Overall (*n* = 1,562)
Mean (SD)	2.2 (0.8)	2.4 (0.9)	2.2 (0.8)

Table A.35. Time to Complete PROMIS Anxiety by Region (minutes)

	Northeast (*n* = 427)	South (*n* = 586)	Midwest (*n* = 302)	West (*n* = 247)	Overall (*n* = 1,562)
Mean (SD)	2.2 (0.8)	2.1 (0.8)	2.1 (0.8)	2.2 (0.9)	2.2 (0.8)

Table A.36. Time to Complete PROMIS Anxiety Data Elements by Facility Ownership (minutes)

	For-Profit (*n* = 951)	Nonprofit (*n* = 597)	Overall (*n* = 1,562[a])
Mean (SD)	2.2 (0.9)	2.2 (0.8)	2.2 (0.8)

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

Table A.37. Time to Complete PROMIS Anxiety Data Elements by Facility Size (minutes)

	Below Setting-Type Median (*n* = 676)	Above Setting-Type Median (*n* = 885)	Overall (*n* = 1,562[a])
Mean (SD)	2.2 (0.9)	2.1 (0.8)	2.2 (0.8)

[a] 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.38. Interrater Reliability Kappa or Weighted Kappa for PROMIS Anxiety items by Urbanicity

Data Element	Urban (*n* = 859)	Nonurban (*n* = 66)
How often past 3/7 days: felt worried (e3a)	0.98	1.00
How often past 3/7 days: overwhelmed by worries (e3b)	0.98	1.00
How often past 3/7 days: trouble paying attention (e3c)	0.97	0.99
How often past 3/7 days: felt nervous (e3d)	0.98	0.99
How often past 3/7 days: difficulty calming down (e3e)	0.97	1.00
How often past 3/7 days: focused on anxiety (e3f)	0.99	1.00
How often past 3/7 days: felt need for help with anxiety (e3g)	0.99	0.98
How often past 3/7 days: sudden feelings of panic (e3h)	0.98	1.00
Anxiety total[a]	0.99	0.99

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

Table A.39. Interrater Reliability Kappa or Weighted Kappa for PROMIS Anxiety items by Region

Data Element	Northeast (*n* = 202)	South (*n* = 349)	Midwest (*n* = 204)	West (*n* = 170)
How often past 3/7 days: felt worried (e3a)	0.99	0.97	0.99	0.99
How often past 3/7 days: overwhelmed by worries (e3b)	0.98	0.97	0.98	0.98
How often past 3/7 days: trouble paying attention (e3c)	0.98	0.97	0.99	0.95
How often past 3/7 days: felt nervous (e3d)	0.97	0.99	0.99	0.97
How often past 3/7 days: difficulty calming down (e3e)	0.97	0.99	0.97	0.95
How often past 3/7 days: focused on anxiety (e3f)	0.98	1.00	0.99	1.00
How often past 3/7 days: felt need for help with anxiety (e3g)	0.98	0.99	0.98	0.99
How often past 3/7 days: sudden feelings of panic (e3h)	0.97	0.99	0.97	1.00
Anxiety total[a]	0.99	0.99	0.99	0.99

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

Table A.40. Interrater Reliability Kappa or Weighted Kappa for PROMIS Anxiety Data Elements by Facility Ownership

Data Element	For-Profit (*n* = 584)	Nonprofit (*n* = 335)
How often past 3/7 days: felt worried (e3a)	0.98	0.99
How often past 3/7 days: overwhelmed by worries (e3b)	0.99	0.96
How often past 3/7 days: trouble paying attention (e3c)	0.98	0.96
How often past 3/7 days: felt nervous (e3d)	0.99	0.97
How often past 3/7 days: difficulty calming down (e3e)	0.99	0.96
How often past 3/7 days: focused on anxiety (e3f)	1.00	0.98
How often past 3/7 days: felt need for help with anxiety (e3g)	0.99	0.98
How often past 3/7 days: sudden feelings of panic (e3h)	1.00	0.96
Anxiety total[a]	1.00	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, 0.81–1.00 is excellent/almost perfect.
[a] Pearson correlation for total scores.

Table A.41. Interrater Reliability Kappa or Weighted Kappa for PROMIS Anxiety Data Elements by Facility Size

Data Element	Below Setting-Type Median (*n* = 414)	Above Setting-Type Median (*n* = 510)
How often past 3/7 days: felt worried (e3a)	0.99	0.98
How often past 3/7 days: overwhelmed by worries (e3b)	0.98	0.98
How often past 3/7 days: trouble paying attention (e3c)	0.98	0.97
How often past 3/7 days: felt nervous (e3d)	0.99	0.98

Data Element	Below Setting-Type Median (n = 414)	Above Setting-Type Median (n = 510)
How often past 3/7 days: difficulty calming down (e3e)	0.98	0.97
How often past 3/7 days: focused on anxiety (e3f)	0.99	0.99
How often past 3/7 days: felt need for help with anxiety (e3g)	0.99	0.98
How often past 3/7 days: sudden feelings of panic (e3h)	0.99	0.98
Anxiety total [a]	1.00	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, 0.81–1.00 is excellent/almost perfect.
[a] Pearson correlation for total scores.

Table A.42. Summed Score to T-Score Conversion Table for Eight PROMIS Anxiety Data Elements Included in National Beta Test

Raw Summed Score	T-Score	Standard Error of T-Score
8	37.1	5.7
9	43.0	3.8
10	45.8	3.3
11	48.0	2.9
12	49.8	2.6
13	51.3	2.5
14	52.6	2.3
15	53.9	2.2
16	55.0	2.2
17	56.1	2.2
18	57.1	2.1
19	58.2	2.1
20	59.2	2.1
21	60.2	2.1
22	61.2	2.1
23	62.3	2.1
24	63.3	2.1
25	64.3	2.2
26	65.4	2.2
27	66.4	2.2
28	67.5	2.2
29	68.5	2.2
30	69.6	2.1
31	70.7	2.2
32	71.8	2.2

Raw Summed Score	T-Score	Standard Error of T-Score
33	72.9	2.2
34	74.1	2.2
35	75.3	2.2
36	76.6	2.3
37	78.0	2.4
38	79.6	2.5
39	81.5	2.7
40	83.9	2.9

NOTE: This table is valid only for the eight items administered in the National Beta Test.

Supplementary Tables for Pain Interview Data Elements

Table A.43. Comparison of Pain Interview Data Elements According to Version

	Market		Statistics			
Market Group A (past 7 days); Market Group B (past 3 days)	**A**	**B**	**Min/max**	**t-value**	**p-value**	**Cohen's *d***
Have you had any pain or hurting any time in the last X days			0/1	3.54	0.00	0.13
Number of assessments	1,520	1,511				
Mean score	0.75	0.80				
Standard deviation	0.43	0.40				
How much of the time have you experienced pain or hurting over the last X days?			1/4	–2.04	0.04	-0.08
Number of assessments	1,140	1,207				
Mean score	2.87	2.80				
Standard deviation	0.89	0.91				
Over the past X days, how much of the time has pain made it hard for you to sleep?			1/4	–2.22	0.03	-0.09
Number of assessments	1,130	1,207				
Mean score	2.21	2.11				
Standard deviation	1.05	1.04				
Over the past X days, have you been offered any rehabilitation therapies (e.g., physical therapy, occupational therapy, speech therapy) by your care providers?			0/1	–1.41	0.16	-0.06
Number of assessments	1,135	1,208				
Mean score	0.90	0.88				
Standard deviation	0.30	0.32				

	Market		Statistics			
Market Group A (past 7 days); Market Group B (past 3 days)	**A**	**B**	**Min/max**	**t-value**	**p-value**	**Cohen's *d***
Over the past X days, how often have you limited your participation in rehabilitation therapy sessions because of pain?			1/4	1.86	0.06	0.08
Number of assessments	1,004	1,037				
Mean score	1.39	1.46				
Standard deviation	0.75	0.82				
Over the past X days, how much of the time have you limited your day-to-day activities (excluding rehabilitation therapy sessions) because of pain?			1/4	–1.16	0.25	-0.05
Number of assessments	1,108	1,190				
Mean score	2.05	1.99				
Standard deviation	1.09	1.09				
Please rate your worst pain over the last X days on a zero-to-ten scale, with zero being no pain and ten being the worst pain you can imagine			0/10	–0.34	0.74	-0.01
Number of assessments	1,128	1,199				
Mean score	7.16	7.13				
Standard deviation	2.49	2.61				
Over the past X days, how much relief have you felt from pain due to pain treatments and/or medications?			0/4	0.26	0.80	0.01
Number of assessments	1,066	1,118				
Mean score	2.63	2.64				
Standard deviation	0.84	0.83				

Table A.44. Frequencies for Frequent or Almost Constant Pain Interference by Patient/Resident Characteristics and Clinical Groups in the HHA Setting (percent)

Patient/Resident Characteristics and Clinical Groups	Any Pain	Interference with Sleep
Gender (*n_any* = 625; *n_int* = 473)		
Male (*n_any* = 224; *n_int* = 161)	71.9	24.4[a]
Female (*n_any* = 401; *n_int* = 312)	77.8	33.9
Age (*n_any* = 621; *n_int* = 470)		
18–44 (*n_any* = 4; *n_int* = 3)	75.0[a]	66.7[a]
45–64 (*n_any* = 60; *n_int* = 52)	86.7	51.0
65–74 (*n_any* = 175; *n_int* = 152)	86.9	36.8
75–89 (*n_any* = 312; *n_int* = 221)	70.8	22.6
90+ (*n_any* = 70; *n_int* = 42)	60.0	25.0

Patient/Resident Characteristics and Clinical Groups	Any Pain	Interference with Sleep
Length of stay (*n_any* = 499; *n_int* = 380; mean, SD)	No: 33.5 (16.1) Yes: 30.3 (15.4)	No: 30.0 (15.6) Yes: 30.9 (15.3)
Disposition at discharge (*n_any* = 617; *n_int* = 469)		
Home (*n_any* = 458; *n_int* = 348)	76.0	29.8
Hospital (*n_any* = 23; *n_int* = 18)	78.3	33.3
Hospice (*n_any* = 12; *n_int* = 9)	75.0	22.2
HHA (*n_any* = 15; *n_int* = 12)	80.0	33.3
IRF (*n_any* = 4; *n_int* = 1)	25.0	0.0
LTCH (*n_any* = 1*; n_int* = 1)	100.0	0.0
SNF (n_any = 6; *n_int* = 5)	83.3	60.0
Other (n_any = 98; *n_int* = 75)	76.5	33.8
Clinical conditions (*n_any* = 420; *n_int* = 320)		
Sepsis		
Yes (*n_any* = 9; *n_int* = 4)	44.4[a]	32.8
No (*n_any* = 411; *n_int* = 316)	76.9	0.0
Heart failure		
Yes (*n_any* = 32; *n_int* = 20)	62.5	10.0[a]
No (*n_any* = 388; *n_int* = 300)	77.3	33.9
Stroke		
Yes (*n_any* = 7; *n_int* = 5)	71.4	0.0
No (*n_any* = 413; *n_int* = 315)	76.3	32.9
Mobility—Lying to sitting (*n_any* = 393; *n_int* = 300)		
Independent (*n_any* = 30; *n_int* = 26)	86.7	30.8
Setup or clean-up (*n_any* = 60; *n_int* = 43)	71.7	32.6
Supervision or touching (*n_any* = 119; *n_int* = 89)	74.8	32.6
Partial/moderate (*n_any* = 125; *n_int* = 101)	80.8	36.0
Substantial/maximal (*n_any* = 53; *n_int* = 38)	71.7	32.4
Dependent (*n_any* = 6; *n_int* = 3)	50.0	0.0

NOTES: *n_any* = number of patients/residents who reported the presence of any pain; *n_int* = number of patients/residents who said that pain interfered with their sleep “frequently” or “almost constantly.” Note that the number for each column may be slightly smaller, and in particular “Interference with Therapy” is further reduced for patients who were not offered therapy.

[a] Significant ($p < 0.05$) associations with interference due to pain.

Table A.45. Frequencies for Pain Data Elements by Patient/Resident Characteristics and Clinical Groups in the IRF Setting (percent)

Patient/Resident Characteristics and Clinical Groups	Any Pain	Interference with Sleep
Gender (*n_any* = 746; *n_int* = 588)		
Male (*n_any* = 319; *n_int* = 235)	73.7[a]	37.8
Female (*n_any* = 427; *n_int* = 353)	82.7	38.5
Age (*n_any* = 743; *n_int* = 586)		
18–44 (*n_any* = 6; *n_int* = 4)	66.7	100[a]
45–64 (*n_any* = 58; *n_int* = 46)	79.3	64.4
65–74 (*n_any* = 289; *n_int* = 234)	81.0	41.0
75–89 (*n_any* = 335; *n_int* = 257)	76.7	30.5
90+ (*n_any* = 55; *n_int* = 45)	81.8	35.6
Length of stay (*n_any* = 727; *n_int* = 570; mean, SD)	No: 14.4 (5.1) Yes: 14.0 (5.1)	No: 13.7 (4.8) Yes: 14.4 (5.4)
Disposition at discharge (*n_any* = 741; *n_int* = 583)		
Home (*n_any* = 318; *n_int* = 238)	74.8	37.8
Hospital (*n_any* = 36; *n_int* = 32)	88.9	37.5
Hospice (*n_any* = 7; *n_int* = 6)	85.7	16.7
HHA (*n_any* = 253; *n_int* = 204)	80.6	37.9
IRF (*n_any* = 1; *n_int* = 1)	100.0	0.0
LTCH (*n_any* = 1; *n_int* = 1)	100.0	100.0
SNF (*n_any* = 105; *n_int* = 86)	81.9	41.2
Other (*n_any* = 20; *n_int* = 15)	75.0	40.0
Clinical conditions (*n_any* = 583; *n_int* = 459)		
Sepsis		
Yes (*n_any* = 25; *n_int* = 22)	88.0	59.1[a]
No (*n_any* = 558; *n_int* = 437)	78.3	36.3
Heart failure		
Yes (*n_any* = 131; *n_int* = 95)	72.5[a]	35.8
No (*n_any* = 452; *n_int* = 364)	80.5	37.9
Stroke		
Yes (*n_any* = 99; *n_int* = 53)	53.5[a]	24.5[a]
No (*n_any* = 484; *n_int* = 406)	83.9	39.1
Hygiene—Toileting (*n_any* = 567; *n_int* = 447)		
Independent (*n_any* = 5; *n_int* = 4)	80.0	25.0
Setup or clean-up (*n_any* = 21; *n_int* = 13)	61.9	23.1
Supervision or touching (*n_any* = 122; *n_int* = 95)	77.9	31.9
Partial/moderate (*n_any* = 136; *n_int* = 105)	77.2	31.4
Substantial/maximal (*n_any* = 139; *n_int* = 111)	79.9	44.1
Dependent (*n_any* = 144; *n_int* = 119)	82.6	41.5

Patient/Resident Characteristics and Clinical Groups	Any Pain	Interference with Sleep
Mobility—Lying to sitting (*n_any* = 580; *n_int* = 456)		
Independent (*n_any* = 42; *n_int* = 28)	66.7[a]	25.0[a]
Setup or clean-up assistance (*n_any* = 16; *n_int* = 10)	62.5	50.0
Supervision or touching (*n_any* = 186; *n_int* = 137)	73.7	32.9
Partial/moderate (*n_any* = 217; *n_int* = 172)	79.3	38.0
Substantial/maximal (*n_any* = 92; *n_int* = 83)	90.2	42.7
Dependent (*n_any* = 27; *n_int* = 26)	96.3	50.0

NOTES: *n_any* = number of patients/residents who reported the presence of any pain; *n_int* = number of patients/residents who said that pain interfered with their sleep "frequently" or "almost constantly." Note that the number for each column may be slightly smaller, and in particular "Interference with Therapy" is further reduced for patients who were not offered therapy.
[a] Significant ($p < 0.05$) associations with interference due to pain.

Table A.46. Frequencies for Pain Data Elements by Patient/Resident Characteristics and Clinical Groups in the LTCH Setting (percent)

Patient/Resident Characteristics and Clinical Groups	Any Pain	Interference with Sleep
Gender (*n_any* = 463; *n_int* = 361)		
Male (*n_any* = 239; *n_int* = 183)	76.6	45.1
Female (*n_any* = 224; *n_int* = 178)	79.5	48.0
Age (*n_any* = 464; *n_int* = 362)		
18–44 (*n_any* = 23; *n_int* = 22)	95.7[a]	45.5
45–64 (*n_any* = 117; *n_int* = 99)	84.6	52.0
65–74 (*n_any* = 165; *n_int* = 136)	82.4	46.7
75–89 (*n_any* = 145; *n_int* = 99)	68.3	42.4
90+ (*n_any* = 14; *n_int* = 6)	42.9	33.3
Length of stay (*n_any* = 409; *n_int* = 322; mean, SD)	No: 20.0 (9.3)[a] Yes: 24.8 (11.3)	No: 24.7 (11.7) Yes: 24.9 (10.8)
Disposition at discharge (*n_any* = 448; *n_int* = 351)		
Home (*n_any* = 91; *n_int* = 66)	72.5[a]	47.0
Hospital (*n_any* = 31; *n_int* = 30)	96.8	40.0
Hospice (*n_any* = 12; *n_int* = 7)	58.3	42.9
HHA (*n_any* = 77; *n_int* = 66)	85.7	43.9
IRF (*n_any* = 45; *n_int* = 38)	84.4	55.3
LTCH (*n_any* = 1; *n_int* = 1)	100.0	100.0
SNF (*n_any* = 128; *n_int* = 102)	79.7	47.0
Other (*n_any* = 63; *n_int* = 41)	65.1	48.8
Clinical conditions (*n_any* = 395; *n_int* = 314)		
Sepsis		
Yes (*n_any* = 67; *n_int* = 53)	79.1	54.7
No (*n_any* = 328; *n_int* = 261)	79.6	45.2

Patient/Resident Characteristics and Clinical Groups	Any Pain	Interference with Sleep
Heart failure		
Yes (*n_any* = 13; *n_int* = 10)	76.9	30.0
No (*n_any* = 382; *n_int* = 304)	79.6	47.4
Stroke		
Yes (*n_any* = 27; *n_int* = 19)	70.4	44.4
No (*n_any* = 368; *n_int* = 295)	80.2	46.9
Hygiene—Toileting (*n_any* = 382; *n_int* = 304)		
Independent (*n_any* = 46; *n_int* = 40)	87.0	47.5
Setup or clean-up assistance (*n_any* = 33; *n_int* = 29)	87.9	37.9
Supervision or touching (*n_any* = 56; *n_int* = 35)	62.5	42.9
Partial/moderate (*n_any* = 51; *n_int* = 40)	78.4	37.5
Substantial/maximal (*n_any* = 63; *n_int* = 52)	82.5	63.5
Dependent (*n_any* = 133; *n_int* = 108)	81.2	48.1
Mobility—Lying to sitting (*n_any* = 343; *n_int* = 274)		
Independent (*n_any* = 63; *n_int* = 53)	84.1	52.8[a]
Setup or clean-up (*n_any* = 24; *n_int* = 20)	83.3	31.6
Supervision or touching (*n_any* = 57; *n_int* = 38)	66.7	29.0
Partial/moderate (*n_any* = 71; *n_int* = 57)	80.3	49.1
Substantial/maximal (*n_any* = 50; *n_int* = 42)	84.0	66.7
Dependent (*n_any* = 78; *n_int* = 64)	82.1	44.4

NOTES: *n_any* = number of patients/residents who reported the presence of any pain; *n_int* = number of patients/residents who said that pain interfered with their sleep "frequently" or "almost constantly." Note that the number for each column may be slightly smaller, and in particular "Interference with Therapy" is further reduced for patients who were not offered therapy.
[a] Significant ($p < 0.05$) associations with interference due to pain.

Table A.47. Frequencies for Pain Data Elements by Patient/Resident Characteristics and Clinical Groups in the SNF Setting (percent)

Patient/Resident Characteristics and Clinical Groups	Any Pain	Interference with Sleep
Gender (*n_any* = 1,092; *n_int* = 848)		
Male (*n_any* = 422; *n_int* = 319)	75.6	37.1
Female (*n_any* = 670; *n_int* = 529)	79.0	34.5
Age (*n_any* = 1,087; *n_int* = 843)		
18–44 (*n_any* = 9; *n_int* = 8)	88.9[a]	37.5[a]
45–64 (*n_any* = 73; *n_int* = 66)	90.4	62.1
65–74 (*n_any* = 286; *n_int* = 245)	85.7	42.2
75–89 (*n_any* = 545; *n_int* = 416)	76.3	31.8
90+ (*n_any* = 174; *n_int* = 108)	12.8	17.8
Length of stay (*n_any* = 945; *n_int* = 738; mean, SD)	No: 20.1 (11.8) Yes: 21.6 (12.3)	No: 22.1 (12.7) Yes: 20.6 (11.6)

Patient/Resident Characteristics and Clinical Groups	Any Pain	Interference with Sleep
Disposition at discharge (*n_any* = 1,070; *n_int* = 831)		
Home (*n_any* = 476; *n_int* = 371)	77.9[a]	36.3
Hospital (*n_any* = 110; *n_int* = 82)	74.6	43.9
Hospice (*n_any* = 9; *n_int* = 6)	66.7	50.0
HHA (*n_any* = 278; *n_int* = 233)	83.8	35.5
IRF (*n_any* = 1; *n_int* = 1)	100.0	0.0
LTCH (*n_any* = 10; *n_int* = 5)	50.0	20.0
SNF (*n_any* = 43; *n_int* = 26)	60.5	20.0
Other (*n_any* = 143; *n_int* = 107)	74.8	26.9
Clinical conditions (*n_any* = 859; *n_int* = 666)		
Sepsis		
Yes (*n_any* = 51; *n_int* = 38)	74.5	42.1
No (*n_any* = 808; *n_int* = 628)	77.7	36.5
Heart failure		
Yes (*n_any* = 209; *n_int* = 153)	73.2	35.5
No (*n_any* = 650; *n_int* = 513)	78.9	37.2
Stroke		
Yes (*n_any* = 64; *n_int* = 51)	79.7	39.2
No (*n_any* = 795; *n_int* = 615)	77.4	36.6
Hygiene—Toileting (*n_any* = 576; *n_int* = 432)		
Independent (*n_any* = 22; *n_int* = 14)	63.6	50.0
Setup or clean-up assistance (*n_any* = 23; *n_int* = 21)	91.3	33.3
Supervision or touching (*n_any* = 143; *n_int* = 101)	70.6	29.7
Partial/moderate (*n_any* = 180; *n_int* = 131)	72.8	37.7
Substantial/maximal (*n_any* = 134; *n_int* = 104)	77.6	34.0
Dependent (*n_any* = 74; *n_int* = 61)	82.4	42.6
Mobility—Lying to sitting (*n_any* = 568; *n_int* = 427)		
Independent (*n_any* = 58; *n_int* = 35)	60.3	51.4
Setup or clean-up (*n_any* = 14; *n_int* = 9)	64.3	44.4
Supervision or touching (*n_any* = 166; *n_int* = 125)	75.3	29.8
Partial/moderate (*n_any* = 206; *n_int* = 161)	78.2	36.2
Substantial/maximal (*n_any* = 98; *n_int* = 78)	79.6	30.8
Dependent (*n_any* = 26; *n_int* = 19)	73.1	52.6

NOTES: *n_any* = number of patients/residents who reported the presence of any pain; *n_int* = number of patients/residents who said that pain interfered with their sleep “frequently” or “almost constantly.” Note that the number for each column may be slightly smaller, and in particular “Interference with Therapy” is further reduced for patients who were not offered therapy.

[a] Significant ($p < 0.05$) associations with interference due to pain.

Table A.48. Time to Complete Pain Data Elements by Urbanicity (minutes)

		Urban	Nonurban	Overall
All	Number of assessments	1,668	109	1,777
	Combined Mean (SD)	2.6 (1.4)	2.9 (1.5)	2.6 (1.4)
No Pain	Number of assessments	418	22	440
	Combined Mean (SD)	1.2 (0.9)	1.0 (0.5)	1.2 (0.9)
Pain	Number of assessments	1,244	87	1,331
	Combined Mean (SD)	3.0 (1.3)	3.4 (1.2)	3.0 (1.3)

Table A.49. Time to Complete Pain Data Elements by Region (minutes)

		Northeast	South	Midwest	West	Overall
All	Number of assessments	479	659	364	275	1,777
	Combined Mean (SD)	2.4(1.4)	2.6(1.5)	2.7(1.4)	2.6(1.4)	2.6(1.4)
No Pain	Number of assessments	139	154	76	71	440
	Combined Mean (SD)	1.1(0.6)	1.1(0.8)	1.5(1.4)	1.1(0.6)	1.2(0.9)
Pain	Number of assessments	336	503	288	204	1,331
	Combined Mean (SD)	3.0(1.2)	3.1(1.3)	3.0(1.3)	3.1 (1.2)	3.0(1.3)

Table A.50. Time to Complete Pain Data Elements by Facility Ownership (minutes)

		For-Profit	Nonprofit	Overall
All	Number of assessments	1,099	663	1,777[a]
	Combined Mean (SD)	2.6 (1.4)	2.6 (1.4)	2.6 (1.4)
No Pain	Number of assessments	267	172	440[a]
	Combined Mean (SD)	1.2 (0.8)	1.2 (1.0)	1.2 (0.9)
Any Pain	Number of assessments	827	490	1,331[a]
	Combined Mean (SD)	3.0 (1.3)	3.1 (1.2)	3.0 (1.3)

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

Table A.51. Time to Complete Pain Data Elements by Facility Size (minutes)

		Below Setting-Type Median	Above Setting-Type Median	Overall
All	Number of assessments	745	1,031	1,777[a]
	Combined Mean (SD)	2.6 (1.4)	2.5 (1.4)	2.6 (1.4)
No Pain	Number of assessments	172	268	440
	Combined Mean (SD)	1.1 (0.6)	1.2 (1.0)	1.2 (0.9)
Any Pain	Number of assessments	572	758	1,331[a]
	Combined Mean (SD)	3.1 (1.3)	3.0 (1.3)	3.0 (1.3)

[a] 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.52. Interrater Reliability Kappa or Weighted Kappa for Pain Data Elements by Urbanicity

Data Element	Urban (*n* = 449)	Rural (*n* = 67)
Any pain (d1)	0.98	1.00
How often experienced pain (d2)	0.97	1.00
How often pain made it hard to sleep (d3)	0.98	1.00
Offered rehab therapies (d4a)	0.95	1.00
How often limited rehab due to pain (d4b)	0.97	1.00
How often limited daily activities due to pain (d4c)	0.98	1.00
Rate worst pain[a] (d5)	0.97	1.00
How much pain relief due to pain treatments and/or medications (d6)	0.97	0.97

NOTE: Interpretation of kappa or weighted kappa is as follows: 0.00–0.20 is slight/poor, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is substantial/good, 0.81–1.00 is excellent/almost perfect.
[a] Pearson correlation for rating of worst pain, which is on a 0–10 scale.

Table A.53. Interrater Reliability Kappa or Weighted Kappa for Pain Data Elements by Region

Data Element	Northeast (*n* = 210)	South (*n* = 358)	Midwest (*n* = 207)	West (*n* = 178)
Any pain (d1)	0.99	0.99	0.96	0.98
How often experienced pain (d2)	0.99	0.97	0.98	0.96
How often pain made it hard to sleep (d3)	0.97	0.97	0.99	0.97
Offered rehab therapies (d4a)	1.00	0.95	0.95	0.95
How often limited rehab due to pain (d4b)	0.96	0.96	0.99	0.97
How often limited daily activities due to pain (d4c)	0.98	0.98	0.99	0.97
Rate worst pain[a] (d5)	0.98	0.97	0.99	0.96
How much pain relief due to pain treatments and/or medications (d6)	0.96	0.98	0.97	0.96

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

[a] Pearson correlation for rating of worst pain, which is on a 0–10 scale.

Table A.54. Interrater Reliability Kappa or Weighted Kappa for Pain Data Elements by Facility Ownership

Data Element	For-Profit (*n* = 601)	Nonprofit (*n* = 346)
Any pain (d1)	0.99	0.98
How often experienced pain (d2)	0.97	0.98
How often pain made it hard to sleep (d3)	0.98	0.97
Offered rehab therapies (d4a)	0.96	-
How often limited rehab due to pain (d4b)	0.97	0.95
How often limited daily activities due to pain (d4c)	0.99	0.98
Rate worst pain[a] (d5)	0.97	0.97
How much pain relief due to pain treatments and/or medications (d6)	0.98	0.95

NOTES: Interrater reliability not shown for data elements with proportions out of range for stable kappa estimate (per study power calculations). Interpretation of kappa or weighted kappa is as follows: 0.00–0.20: slight/poor, 0.21–0.40: fair, 0.41–0.60: moderate, 0.61–0.80: substantial/good, 0.81–1.00: excellent/almost perfect.
[a] Pearson correlation for rating of worst pain, which is on a 0–10 scale.

Table A.55. Interrater Reliability Kappa or Weighted Kappa for Pain Data Elements by Facility Size

Data Element	Below Setting-Type Median (*n* = 423)	Above Setting-Type Median (*n* = 529)
Any pain (d1)	0.97	0.99
How often experienced pain (d2)	0.99	0.96
How often pain made it hard to sleep (d3)	0.99	0.97
Offered rehab therapies (d4a)	0.98	-
How often limited rehab due to pain (d4b)	0.97	0.97
How often limited daily activities due to pain (d4c)	0.99	0.97
Rate worst pain[a] (d5)	1.00	0.96
How much pain relief due to pain treatments and/or medications (d6)	0.98	0.97

NOTES: Interrater reliability not shown for data elements with proportions out of range for stable kappa estimate (per study power calculations). Interpretation of kappa or weighted kappa is as follows: 0.00–0.20: slight/poor, 0.21–0.40: fair, 0.41–0.60: moderate, 0.61–0.80: substantial/good, 0.81–1.00: excellent/almost perfect.
[a] Pearson correlation for rating of worst pain, which is on a 0–10 scale.

References

Abdulla, Aza, Nicola Adams, Margaret Bone, Allison 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, 2013, pp. i1–i57.

Alexander, Beth J., Paula Plank, Mary B. Carlson, Peggy Hanson, Kristi Picken, and Kay Schwebke, "Methods of Pain Assessment in Residents of Long-Term Care Facilities: A Pilot Study," *Journal of the American Medical Directors Association*, Vol. 6, No. 2, 2005, pp. 137–143.

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, 2009, pp. 1331–1346.

American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*, Washington, D.C.: American Psychiatric Association Publishing, 2013.

Ayerbe, Luis, Salma Ayis, Charles D. A. Wolfe, and Anthony G. Rudd, "Natural History, Predictors and Outcomes of Depression After Stroke: Systematic Review and Meta-Analysis," *British Journal of Psychiatry*, Vol. 202, No, 1, 2013, pp. 14–21.

Ayers, Catherine R., John T. Sorrell, Steven R. Thorp, and Julie Loebach Wetherell, "Evidence-Based Psychological Treatments for Late-Life Anxiety," *Psychology and Aging*, Vol. 22, No. 1, 2007, pp. 8–17.

Bahreini, Maryam, Mohammad Jalili, and Maziar Moradi-Lakeh, "A Comparison of Three Self-Report Pain Scales in Adults with Acute Pain," *Journal of Emergency Medicine*, Vol. 48, No. 1, January 2015, pp. 10–18.

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, Daniel G., *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, 2018, pp. 663–670.

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

Brown, Ellen L., Martha L. Bruce, Gail. J. McAvay, Patrick J. Raue, Mark S. Lachs, and Pamela Nassisi, "Recognition of Late-Life Depression in Home Care: Accuracy of the Outcome and Assessment Information Set," *Journal of the American Geriatrics Society*, Vol. 52, No. 6, 2004, pp. 995–999.

Brown, Mary N., Kate L. Lapane, and Andrea F. Luisi, "The Management of Depression in Older Nursing Home Residents," *Journal of the American Geriatrics Society*, Vol. 50, No. 1, 2002, pp. 69–76.

Bruce, Martha L., Gail J. McAvay, Patrick J. Raue, Ellen L. Brown, Barnett S. Meyers, Dennis J. Keohane, David R. Jagoda, and Carol Weber, "Major Depression in Elderly Home Health Care Patients," *American Journal of Psychiatry*, Vol. 159, No. 8, 2002, pp. 1367–1374.

Burton, C. Alexia Campbell, Jenni Murray, John Holmes, Felicity Astin, Darren Greenwood, and Peter Knapp, "Frequency of Anxiety After Stroke: A Systematic Review and Meta-Analysis of Observational Studies," *International Journal of Stroke*, Vol. 8, No. 7, 2013, pp. 545–559.

Carrière, Isabelle, Joanne Ryan, Joanna Norton, Jacqueline Scali, Robert Stewart, Karen Ritchie, and Marie Laure Ancelin, "Anxiety and Mortality Risk in Community-Dwelling Elderly People," *British Journal of Psychiatry*, Vol. 203, No. 4, 2013, pp. 303–309.

Carson, A. J., K. Postma, J. Stone, C. Warlow, and M. Sharpe, "The Outcome of Depressive Disorders in Neurology Patients: A Prospective Cohort Study," *Journal of Neurology, Neurosurgery & Psychiatry*, Vol. 74, No. 7, 2003, pp. 893–896.

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, 2003, pp. 664–672.

Chin, Raymond P., Chin-Hung Ho, and Lydia Cheung, "Scheduled Analgesic Regimen Improves Rehabilitation After Hip Fracture Surgery," *Clinical Orthopaedics and Related Research*, Vol. 471, No. 7, 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.

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.

Dewitt, Barry, David Feeny, Baruch Fischhoff, David Cella, Ron D. Hays, Rachel Hess, Paul A. Pilkonis, Dennis A. Revicki, Mark S. Roberts, Joel Tsevat, Lan Yu, and Janel Hanmer, "Estimation of a Preference-Based Summary Score for the Patient-Reported Outcomes Measurement Information System: The PROMIS®-Preference (PROPr) Scoring System," *Medical Decision Making*, Vol. 38, No.6, 2018, pp. 683–698.

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, 2012, pp. 828–835.

Diefenbach, Gretchen J., and John Goethe, "Clinical Interventions for Late-Life Anxious Depression," *Clinical Interventions in Aging*, Vol. 1, No. 1, 2006, pp. 41–50.

DiMatteo, M. Robin, Heidi S. Lepper, and Thomas W. Croghan, "Depression Is a Risk Factor for Noncompliance with Medical Treatment: Meta-Analysis of the Effects of Anxiety and Depression on Patient Adherence," *Archives of Internal Medicine*, Vol. 160, No. 14, 2000, pp. 2101–2107.

Djernes, J. K., "Prevalence and Predictors of Depression in Populations of Elderly: A Review," *Acta Psychiatrica Scandinavica*, Vol. 113, No. 5, 2006, pp. 372–387.

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, June 2018. As of June 4, 2019: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/Alpha-2-SPADE-Pilot-Summary-Document.pdf

Edelen, Maria Orlando, Anthony Rodriguez, Sangeeta C. Ahluwalia, Emily K. Chen, 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 15, 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 15, 2019: https://www.rand.org/pubs/research_reports/RR3004z2.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 8: Observational Assessments of Cognitive Function, Mental Status, and Pain)*, Santa Monica, Calif.: RAND Corporation, RR-3004/7-CMS, 2019c As of August 20, 2019: https://www.rand.org/pubs/research_reports/RR3004z7.html

Ell, Kathleen, Jurgen Unützer, Maria Aranda, Kathleen Sanchez, and Pay-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, 2016, pp. 450–457.

Fann, Jesse, Charles Bombardier, Surreya Dikmen, Peter Esselman, Catherine Warms, Erika Pelzer, Holly Rau, and Nancy Temkin, "Validity of the Patient Health Questionnaire-9 in Assessing Depression Following Traumatic Brain Injury," *Journal of Head Trauma Rehabilitation*, Vol. 20, No. 6*,* 2005, pp. 501–511.

Feher, Edward P., Glenn J. Larrabee, and Thomas H. Crook III, "Factors Attenuating the Validity of the Geriatric Depression Scale in a Dementia Population," *Journal of the American Geriatrics Society*, Vol. 40, No. 9, 1992, pp. 906–909.

Ferrell, Bruce A., Wendy M. Stein, and John C. Beck, "The Geriatric Pain Measure: Validity, Reliability and Factor Analysis," *Journal of the American Geriatrics Society*, Vol. 48, No. 12, 2000, pp. 1669–1673.

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, 2011, pp. 288–297.

Gilbody, Simon, David Richards, Stephen Brealey, and Catherine Hewitt, "Screening for Depression in Medical Settings with the Patient Health Questionnaire (PHQ): A Diagnostic Meta-Analysis," *Journal of General Internal Medicine*, Vol. 22, No. 11, 2007, pp. 1596–1602.

Givens, Jane L., Richard N. Jones, and Sharon K. Inouye, "The Overlap Syndrome of Depression and Delirium in Older Hospitalized Patients," *Journal of the American Geriatrics Society*, Vol. 57, No. 8, 2009, pp. 1347–1353.

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

Gloth, F. M., III, A. A. Scheve, C. V. Stober, Selina Chow, and Jane Prosser, "The *Functional Pain Scale*: Reliability, Validity, and Responsiveness in an Elderly Population," *Journal of the American Medical Directors Association*, Vol. 2, No. 3, 2001, pp. 110–114.

Granja, Cristina, Cláudia Dias, Altamiro Costa-Pereira, and António Sarmento, "Quality of Life of Survivors from Severe Sepsis and Septic Shock May Be Similar to That of Others Who Survive Critical Illness," *Critical Care*, Vol. 8, 2004, pp. R91–R98.

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, 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, 2010, pp. 260–269.

Herr, Keela, Karen Bjoro, and Sheila Decker, "Tools for Assessment of Pain in Nonverbal Older Adults with Dementia: A State-of-the-Science Review," *Journal of Pain and Symptom Management*, Vol. 31, No. 2, February 2006, pp. 170–192.

Heyland, Daren, Wilma Hopman, Helen Coo, Joan Tranmer, and Mary McColl, "Long-Term Health-Related Quality of Life in Survivors of Sepsis. Short Form 36: A Valid and Reliable

Measure of Health-Related Quality of Life," *Critical Care Medicine*, Vol. 28, No. 11, November 2000, pp. 3599–3605.

Hjermstad, Marianne Jensen, Peter M. Fayers, Dagny F. Haugen, Augusto Caraceni, Geoffrey W. Hanks, Jon H. Loge, Robin Fainsinger, Nina Aass, Stein Kassam, and European Palliative Care Research Collaborative, "Studies Comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for Assessment of Pain Intensity in Adults: A Systematic Literature Review," *Journal of Pain and Symptom Management*, Vol. 41, No. 6, June 2011, pp. 1073–1093.

Hunnicutt, Jacob N., Christine M. Ulbricht, Jennifer Tjia, and Kate L. Lapane, "Pain and Pharmacologic Pain Management in Long-Stay Nursing Home Residents," *Pain*, Vol. 158, No. 6, 2017, pp. 1091–1099.

Hyer, Lee, "Depression in Long-Term Care," *Clinical Psychology: Science and Practice*, Vol. 12, No. 3, 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, 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, 2012, pp. 11–23.

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, 2003, pp. 1404–1409.

Jongenelis, K., A. M. Pot, A. M. H. Eisses, A. T. F. Beekman, H. Kluiter, and M. W. Ribbe, "Prevalence and Risk Indicators of Depression in Elderly Nursing Home Patients: The AGED Study," *Journal of Affective Disorders*, Vol. 83, Nos. 2–3, 2004, pp. 135–142.

Kamholz, Barbara, and Jürgen Unützer, "Depression After Hip Fracture," *Journal of the American Geriatrics Society*, Vol. 55, No. 1, January 2007, pp. 126–127.

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, 2010, pp. 446–452.

Katon, Wayne, Michael Von Korff, Paul Ciechanowski, Joan Russo, Elizabeth Lin, Gregory Simon, Evette Ludman, Edward Walker, Terry Bush, and Bessie Young "Behavioral and Clinical Factors Associated with Depression Among Individuals with Diabetes," *Diabetes Care*, Vol. 27, No. 4, 2004, pp. 914–920.

Kaup, Bruce A., David Loreck, Ann L. Gruber-Baldini, Pearl German, A. S. Menon, Sheryl Zimmerman, Lynda Burton, and Jay Magaziner, "Depression and Its Relationship to Function and Medical Status, by Dementia Status, in Nursing Home Admissions," *American Journal of Geriatric Psychiatry*, Vol. 15. No. 5, 2007, pp. 438–442.

Keller, San, Carla Bann, Sheri Dodd, Jeff Schein, Tito Mendoza, and Charles Cleeland, "Validity of the Brief Pain Inventory for Use in Documenting the Outcomes of Patients with Noncancer Pain," *Clinical Journal of Pain*, Vol. 20, No. 5, 2004, pp. 309–318.

Kessler, Ronald C., Wai Tat Chiu, Olga Demler, and Ellen E. Walters, "Prevalence, Severity, and Comorbidity of 12-Month DSM-IV Disorders in the National Comorbidity Survey Replication," *Archives of General Psychiatry*, Vol. 62, No. 6, 2005, pp. 617–627.

Kessler, Ronald C., Maria Petukhova, Nancy A. Sampson, Alan M. Zaslavsky, and Hans-Ullrich Wittchen, "Twelve-Month and Lifetime Prevalence and Lifetime Morbid Risk of Anxiety and Mood Disorders in the United States," *International Journal of Methods in Psychiatric Research*, Vol. 21, No. 3, 2012, pp. 169–184.

Klapow, Joshua, Kurt Kroenke, Trudi Horton, Steven Schmidt, Robert Spitzer, and Janet B. Williams, "Psychological Disorders and Distress in Older Primary Care Patients: A Comparison of Older and Younger Samples," *Psychosomatic Medicine*, Vol. 64, No. 4, 2002, pp. 635–643.

Kroenke, Kurt, Robert Spitzer, and Janet B. Williams, "The PHQ-9: Validity of a Brief Depression Severity Measure," *Journal of General Internal Medicine*, Vol. 16, No. 9, 2001, pp. 606–613.

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, 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, 2012 pp. 344–349.

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, 2009, pp. 67–73.

Li, Chunyu, Bruce Friedman, Yeates Conwell, and Kevin Fiscella, "Validity of the Patient Health Questionnaire 2 (PHQ-2) in Identifying Major Depression in Older People," *Journal of the American Geriatrics Society*, Vol. 55, No. 4, 2007, pp. 596–602.

Lichtner, Valentina, Dawn Dowding, Philip Esterhuizen, S. José Closs, Andrew F. Long, Anne Corbett, and Michelle Briggs, "Pain Assessment for People with Dementia: A Systematic Review of Systematic Reviews of Pain Assessment Tools," *BMC Geriatrics*, Vol. 14, Article No. 138, 2014.

Löwe, Bernd, Kurt Kroenke, and Kerstin Gräfe, "Detecting and Monitoring Depression with a Two-Item Questionnaire (PHQ-2)," *Journal of Psychosomatic Research*, Vol. 58, No, 2, 2005, pp. 163–171.

Löwe, Bernd, Kurt Kroenke, Wolfgang Herzog, and Kerstin Gräfe, "Measuring Depression Outcome with a Brief Self-Report Instrument: Sensitivity to Change of the Patient Health Questionnaire (PHQ-9)," *Journal of Affective Disorders*, Vol. 81, No. 1, 2004, pp. 61–66.

Lukas, A., J. B. Barber, P. Johnson, and S. J. Gibson, "Observer-Rated Pain Assessment Instruments Improve Both the Detection of Pain and the Evaluation of Pain Intensity in People with Dementia," *European Journal of Pain*, Vol. 17, No. 10, 2013, pp. 1558–1568.

Luppa, Melanie, Tobias Luck, Herbert Matschinger, Hans-Helmut König, and Steffi G. Riedel-Heller, "Predictors of Nursing Home Admission of Individuals Without a Dementia Diagnosis Before Admission—Results from the Leipzig Longitudinal Study of the Aged (LEILA 75+)," *BMC Health Services Research*, Vol. 10, Article No. 186, 2010.

Manea, Laura, Simon Gilbody, and Dean McMillan, "Optimal Cut-Off Score for Diagnosing Depression with the Patient Health Questionnaire (PHQ-9): A Meta-Analysis," *Canadian Medical Association Journal*, Vol. 184, No. 3, 2012, pp. E191–E196.

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, 1959, pp. 719–748.

McLean, Carmen P., Anu Asnaani, Brett T. Litz, and Stefan G. Hofmann, "Gender Differences in Anxiety Disorders: Prevalence, Course of Illness, Comorbidity and Burden of Illness," *Journal of Psychiatric Research*, Vol. 45, No. 8, 2011, pp. 1027–1035.

Moriarty, Andrew Stephen, Simon Gilbody, Dean McMillan, and Laura Manea, "Screening and Case Finding for Major Depressive Disorder Using the Patient Health Questionnaire (PHQ-9): A Meta-Analysis," *General Hospital Psychiatry*, Vol. 37, No. 6, 2015, pp. 567–576.

Morley, John E., "Weight Loss in the Nursing Home," *Journal of the American Medical Directors Association*, Vol. 8, No. 4, 2007, pp. 201–204.

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

Ocañez, Kendra L. S., R. Kathryn McHugh, and Michael W. Otto, "A Meta-Analytic Review of the Association Between Anxiety Sensitivity and Pain," *Depression and Anxiety*, Vol. 27, No. 8, 2010, pp. 760–767.

Osborn, Amanda Jane, Jane L. Mathias, and A. Kate Fairweather-Schmidt, "Prevalence of Anxiety Following Adult Traumatic Brain Injury: A Meta-Analysis Comparing Measures, Samples and Postinjury Intervals," *Neuropsychology*, Vol. 30, No. 2, 2016, pp. 247–261.

Payne, J. L., J. M. Sheppard, M. Steinberg, A. Warren, A. Baker, C. Steele, J. Brandt, and C. 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, 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.

Pilkonis, Paul A., Seung W. Choi, Steven P. Reise, Angela M. Stover, William T. Riley, and David Cella, "Item Banks for Measuring Emotional Distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): Depression, Anxiety, and Anger," *Assessment*, Vol. 18, No. 3, 2011, pp. 263–283.

Préville, M., G. Côté, R. Boyer, and R. Hébert, "Detection of Depression and Anxiety Disorders by Home Care Nurses," *Aging & Mental Health*, Vol. 8, No. 5, 2004, pp. 400–409.

PROMIS—*See* Barry Dewitt, David Feeny, Baruch Fischhoff, David Cella, Ron D. Hays, Rachel Hess, Paul A. Pilkonis, Dennis A. Revicki, Mark S. Roberts, Joel Tsevat, Lan Yu, and Janel Hanmer, "Estimation of a Preference-Based Summary Score for the Patient-Reported Outcomes Measurement Information System: The PROMIS®-Preference (PROPr) Scoring System," *Medical Decision Making*, Vol. 38, No.6, 2018, pp. 683–698.

Radloff, Lenore Sawyer, "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population," *Applied Psychological Measurement*, Vol. 1, No. 3, 1977, pp. 385–401.

Rahman, Anna N., and Robert A. Applebaum, "The Nursing Home Minimum Data Set Assessment Instrument: Manifest Functions and Unintended Consequences—Past, Present, and Future," *Gerontologist*, Vol. 49, No. 6, 2009, pp. 727–735.

RAND Corporation, *Technical Expert Panel Summary/Expert Input Report: Development and Maintenance of Post-Acute Care Cross-Setting Standardized Patient Assessment Data*, Santa Monica, Calif., WR-1187-CMMS, 2017a. As of June 28, 2019: https://www.rand.org/pubs/working_papers/WR1187.html

———, *Technical Expert Panel Summary (Second Convening): Development and Maintenance of Post-Acute Care Cross-Setting Standardized Patient Assessment Data*, Santa Monica,

Calif., WR-1184-CMMS, 2017b. As of June 28, 2019: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/RAND-IMPACT-TEP-Second-Convening-Final-Report-March-2017.pdf

Rapp, Michael A., Michael Schnaider-Beeri, Michael Wysocki, Elizabeth Guerrero-Berroa, Hillel T. Grossman, Andreas Heinz, and Vahram Haroutunian, "Cognitive Decline in Patients with Dementia as a Function of Depression," *American Journal of Geriatric Psychiatry*, Vol. 19, No. 4, 2011, pp. 357–363.

Reise, Steven P., "The Rediscovery of Bifactor Measurement Models," *Multivariate Behavioral Research*, Vol. 47, No. 5, 2012, pp. 667–696.

Reynolds, William M., and Kenneth A. Kobak, "Reliability and Validity of the Hamilton Depression Inventory: A Paper-and-Pencil Version of the Hamilton Depression Rating Scale Clinical Interview," *Psychological Assessment*, Vol. 7, No. 4, 1995, pp. 472–483.

Robinson, Robert G., "Poststroke Depression: Prevalence, Diagnosis, Treatment, and Disease Progression," *Biological Psychiatry*, Vol. 54, No. 3, 2003, pp. 376–387.

Rodriguez, Anthony, Steven P. Reise, and Mark G. Haviland, "Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures," *Journal of Personality Assessment*, Vol. 98, No. 3, 2016a, pp. 223–237.

———, "Evaluating Bifactor Models: Calculating and Interpreting Statistical Indices," *Psychological Methods*, Vol. 21, No. 2, 2016b, pp. 137–150.

Rohan, Kelly J., Jennifer N. Rough, Maggie Evans, Sheau-Yan Ho, Jonah Meyerhoff, Lorinda M. Roberts, and Pamela M. Vacek, "A Protocol for the Hamilton Rating Scale for Depression: Item Scoring Rules, Rater Training, and Outcome Accuracy with Data on Its Application in a Clinical Trial," *Journal of Affective Disorders*, Vol. 200, 2016, pp. 111–118.

Roth, David L., Michelle L. Ackerman, Ozioma C. Okonkwo, and Louis D. Burgio, "The Four-Factor Model of Depressive Symptoms in Dementia Caregivers: A Structural Equation Model of Ethnic Differences," *Psychology and Aging*, Vol. 23, No. 3, 2008, p. 567–576.

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

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, 2012, pp. 618–625.

Samus, Quincy M., Lawrence Mayer, Chiadi U. Onyike, Jason Brandt, Alva Baker, Matthew McNabney, Peter V. Rabins, Constantine G. Lyketsos, and Adam Rosenblatt, "Correlates of Functional Dependence Among Recently Admitted Assisted Living Residents with and Without Dementia," *Journal of the American Medical Directors Association*, Vol. 10, No. 5, 2009, pp. 323–329.

Schalet, Benjamin D., Paul A. Pilkonis, Lan Yu, Nathan Dodds, Kelly L. Johnston, Susan Yount, William Riley, and David Cella, "Clinical Validity of PROMIS Depression, Anxiety, and Anger Across Diverse Clinical Samples," *Journal of Clinical Epidemiology*, Vol. 73, May 2016, pp. 119–127.

Schoevers, R. A., A. T. F. Beekman, D. J. H. Deeg, C. Jonker, and W. van Tilburg, "Comorbidity and Risk-Patterns of Depression, Generalised Anxiety Disorder and Mixed Anxiety-Depression in Later Life: Results from the AMSTEL Study," *Geriatric Psychiatry*, Vol. 18, No. 11, November 2003, pp. 994–1001.

Schultz, Susan K., Angela Hoth, and Kathleen Buckwalter, "Anxiety and Impaired Social Function in the Elderly," *Annals of Clinical Psychiatry*, Vol. 16, No. 1, 2004, pp. 47–51.

Seitz, Dallas, Nitin Purandare, and David Conn, "Prevalence of Psychiatric Disorders Among Older Adults in Long-Term Care Homes: A Systematic Review," *International Psychogeriatrics*, Vol. 22, No. 7, 2010, pp. 1025–1039.

Shao, Huibo, Timonthy R. Peng, Martha L. Bruce, and Yuhua Bao, "Diagnosed Depression Among Medicare Home Health Patients: National Prevalence Estimates and Key Characteristics," *Psychiatric Services*, Vol. 62, No. 5, 2011, pp. 538–540.

Sheeran, Thomas, Amy L. Byers, and Martha L. Bruce, "Depression and Increased Short-Term Hospitalization Risk Among Geriatric Patients Receiving Home Health Care Services," *Psychiatric Services*, Vol. 61, No. 1, January 2010, pp. 78–80.

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 (Biological Sciences and Medical Sciences)*, Vol. 70, No. 5, January 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, 2011, pp. 54–64.

Smalbrugge, M., A. M. Pot, K. Jongenelis, A. T. F. Beekman, and J. A. Eefsting, "Prevalence and Correlates of Anxiety Among Nursing Home Patients," *Journal of Affective Disorders*, Vol. 88, No. 2, 2005, pp. 145–153.

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.

Stein, Murray, Peter Roy-Byrne, Michelle Craske, Alexander Bystritsky, Greer Sullivan, Jeffrey Pyne, Wayne Katon, and Cathy Sherbourne, "Functional Impact and Health Utility of Anxiety Disorders in Primary Care Outpatients," *Medical Care*, Vol. 43, No. 12, December 2005, pp. 1164–1170.

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, 2014, pp. 1100–1105.

Svanborg, Pär, and Marie Åsberg, "A Comparison Between the Beck Depression Inventory (BDI) and the Self-Rated Version of the Montgomery-Åsberg Depression Rating Scale (MADRS)," *Journal of Affective Disorders*, Vol. 64, Nos. 2–3, 2001, pp. 203–216.

Tan, Gabriel, Mark. P. Jensen, John I. Thornby, and Bilal F. Shanti, "Validation of the Brief Pain Inventory for Chronic Nonmalignant Pain," *Journal of Pain*, Vol. 5, No. 12, 2004, pp. 133–137.

Teno, Joan M., Glen Kabumoto, Terrie Wetle, Jason Roy, and Vincent Mor, "Daily Pain That Was Excruciating at Some Time in the Previous Week: Prevalence, Characteristics, and Outcomes in Nursing Home Residents," *Journal of the American Geriatrics Society*, Vol. 52, No. 5, 2004, pp. 762–767.

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

U.S. Department of Agriculture, "Documentation: 2010 Rural-Urban Commuting Area (RUCA) Codes," webpage, last updated October 12, 2016. As of June 21, 2019: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/documentation/

Williams, Linda S., Edward J. Brizendine, Laurie Plue, Tamilyn Bakas, Wanzhu Tu, Hugh Hendrie, and Kurt Kroenke, "Performance of the PHQ-9 as a Screening Tool for Depression After Stroke," *Stroke*, Vol. 36, No. 3, 2005, pp. 635–638.

Wittkampf, Karin A., Leonie Naeije, Aart H. Schene, Jochanan Huyser, and Henk C. van Weert, "Diagnostic Accuracy of the Mood Module of the Patient Health Questionnaire: A Systematic Review," *General Hospital Psychiatry*, Vol. 29, No. 5, 2007, pp. 388–395.

Won, Aida, Kate Lapane, Giovanni Gambassi, Roberto Bernabei, Vince Mor, and Lewis A. Lipsitz, "Correlates and Management of Nonmalignant Pain in the Nursing Home," *Journal of the American Geriatrics Society*, Vol. 47, No. 8, 1999, pp. 936–942.

Yesavage, Jerome A., and Javaid I. Sheikh, "9/Geriatric Depression Scale (GDS): Recent Evidence and Development of a Shorter Version," *Clinical Gerontologist*, Vol. 5, Nos. 1–2, 1986, pp. 165–173.

Young, Yuchi, "Factors Associated with Permanent Transition from Independent Living to Nursing Home in a Continuing Care Retirement Community," *Journal of the American Medical Directors Association*, Vol. 10, No. 7, 2009, pp. 491–497.

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, 2013, pp. S137–S144.

Ziegelstein, Roy C., "Depression in Patients Recovering from a Myocardial Infarction," *Journal of the American Medical Association*, Vol. 286, No. 13, 2001, pp. 1621–1627.

Zwakhalen, Sandra M. G., Jan P. H. Hamers, Huda Huijer Abu-Saad, and Martijn P. F. Berger, "Pain in Elderly People with Severe Dementia: A Systematic Review of Behavioural Pain Assessment Tools," *BMC Geriatrics*, Vol. 6, Article No. 3, 2006.