

Development and Evaluation of Candidate Standardized Patient Assessment Data Elements: Findings from the National Beta Test (Volume 1: Executive Summary)

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-3004/1-CMS



**Development and Evaluation of Candidate Standardized
Patient Assessment Data Elements:
Findings from the National Beta Test
(Volume 1: Executive Summary)**

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/RR3004z1

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.

Published in 2019

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).

Research Team

Project Leadership

Maria Edelen, Ph.D. RAND Health	Project Director
Emily Chen, Ph.D. RAND Health	Project Co-Director
Susan Paddock, Ph.D. RAND Health	Lead Statistician
Anthony Rodriguez, Ph.D. RAND Health	Lead Psychometrician
Sangeeta Ahluwalia, Ph.D.	Task Lead for Training
Rosa Elena Garcia, M.P.H.	Survey Director
Debra Saliba, M.D., M.P.H.	Senior Clinical Advisor

RAND Project Staff

Emily Butcher, B.A.
Catherine Cohen, Ph.D.
Sarah Dalton, M.A.
Michael Dunbar, Ph.D.
Jason Etchegaray, Ph.D.
Shira Fischer, M.D., Ph.D.
Liisa Hiatt, M.S.
Wenjing Huang, Ph.D.
David Klein, M.S.
Jaime Madrigano, Sc.D.
Monique Martineau, M.A.
Patrick Orr, B.A.
Elizabeth Petrun-Sayers, Ph.D.
Lauren Pfeifer, B.A.
Jessica Phillips, M.S.
Francesca Pillemer, Ph.D.
Nabeel Qureshi, M.P.H.
Rachel Reid, M.D.
Rachel Ross, M.P.H.
Shoshana Shelton, M.P.H.
Cathy Sherbourne, Ph.D.

Victoria Shier, Ph.D.
Regina Shih, Ph.D.
Molly Simmons, Ph.D.
Elizabeth Sloss, Ph.D.

Abt Project Staff

Terry Moore, M.P.H., B.S.N.	Principal Associate/V.P., Health Policy
Teresa Mota, B.S.N., R.N.	Nurse Researcher/Associate
Brenda Karkos, M.S.N./M.B.A., R.N.	Associate-Nurse Researcher
Lauren Christopher	Research Assistant
Claire Hoffman	Associate Analyst
Olivia Jung	Analyst
Kelly Weiss	Research Assistant

Qualidigm Project Staff

Ann Spenard, M.S.N., R.N.-B.C.	Vice President/Principal
Lisa Newton, M.S.N., R.N.-B.C.	Consultant
Julia Portale, M.B.A., M.P.H.	Consulting Director

Atlas Project Staff

Julia Rollison, Ph.D., M.P.H.	Senior Principal
Jessica Danaux	Consultant
Lauren Whiteman	Consultant

Northwestern Project Staff

Benjamin David Schalet, Ph.D.	Director, subcontract
Aaron James Kaat, Ph.D.	Psychometrician
John Devin Peipert, Ph.D.	Psychometrician
Christa Martens, M.A.	Project Manager

Contents

Preface.....	iii
Research Team.....	iv
Tables.....	vii
Acknowledgments.....	viii
Abbreviations.....	ix
1. Background.....	1
The Case for PAC Standardized Patient Assessment.....	1
Project Overview.....	2
Stakeholder Input Opportunities.....	2
Candidate SPADEs in the National Beta Test.....	5
2. Methods.....	9
Design of Data Collection.....	9
Market and Facility Sampling and Recruitment.....	10
Assessor Training.....	10
Data Collection.....	11
Analytic Plan.....	12
Sample.....	13
3. Results.....	16
Results for Data Elements Related to Cognitive Function.....	16
Results for Data Elements Related to Mental Status.....	21
Results for Data Elements Related to Special Services, Treatments, and Interventions.....	24
Results for Data Elements Related to Medical Conditions: Pain.....	27
Results for Data Elements Related to Impairments.....	28
Results for Data Elements Related to Other Clinical Categories.....	32
Results for Data Elements Related to Observational Assessments of Cognitive Function, Mental Status, and Pain.....	36
Additional Results Related to Evaluation of SPADE Validity.....	39
4. Discussion.....	41
Summary Evaluation of Candidate SPADEs.....	42
References.....	48

Tables

Table 2.1. Total and Average Assessment Counts by National Beta Test Subsample, Overall and by PAC Setting Type	14
Table 4.1 Summary of Performance of Data Elements Tested in the National Beta Test.....	46

Acknowledgments

We wish to acknowledge the insightful guidance and input received from staff at the Centers for Medicare & Medicaid Services, including Stella Mandl, Mary Pratt, Tara McMullen, Teresa Mota, and Charlayne Van.

We are grateful for the contributions of the RAND, Abt, and Qualidigm team members listed in the “Research Team” section. We thank all of the provider organizations that contributed their staff, their time, and their energy to making the pilot and national field tests a success. A special thanks also to the thousands of patients and residents and their families who participated in the pilot and national field tests.

We extend our appreciation to the wide variety of stakeholders, especially to the advisers and technical expert panel members, who provided input and perspectives that guided selection of the data elements to include in the field test. Finally, we thank Justin Timbie of RAND and Barbara Gage of George Washington University for their thoughtful reviews of the report.

Abbreviations

BIMS	Brief Interview for Mental Status
CARE	Continuity Assessment Record and Evaluation
CMS	Centers for Medicare & Medicaid Services
CY	calendar year
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
IV	intravenous
LCDS	Long-Term Care Hospital CARE Data Set
LTCH	long-term care hospital
MDS	Minimum Data Set
OASIS	Outcome and Assessment Information Set
PAC	post-acute care
PC	public comment
PHQ	Patient Health Questionnaire
PROMIS	Patient Reported Outcome Measurement Information System
SNF	skilled nursing facility
SPADE	standardized patient assessment data element
SSTIs	special services, treatments and interventions
TEP	technical expert panel

1. Background

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).

The Case for PAC Standardized Patient Assessment

Implementation of SPADEs has important implications for Medicare beneficiaries, families, providers, and policymakers. Forty percent of fee-for-service Medicare beneficiaries discharged from an acute care hospital are admitted to a PAC setting.¹ Using SPADEs across PAC settings offers the potential to improve coordination and integration of care, move toward more patient-centered care, improve discharge, facilitate interoperability, and develop outcomes-focused value-based payment models.

Patients who receive care in PAC settings are assessed with a variety of instruments, including the CMS assessment instruments. The 46 percent of PAC patients/residents receiving care in SNFs are assessed with the Minimum Data Set (MDS) for nursing homes; 44 percent of PAC patients/residents are receiving care from HHAs, for which the assessment instrument is the Outcome and Assessment Information Set (OASIS). Eight percent of PAC patients/residents are receiving care in IRFs, for which the assessment instrument is the Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI). About 2 percent of PAC patients/residents are receiving care in LTCHs, for which the assessment instrument is the Long-Term Care Hospital Continuity Assessment Record and Evaluation (CARE) Data Set (LCDS).²

Data from these assessments are used for payment, survey, quality measurement and improvement, and care planning.³ Although these assessment instruments measure similar clinical concepts and similar patients can be treated in multiple PAC settings, each instrument uses different assessment data elements. Medicare uses different payment models and varying

¹ Medicare Payment Advisory Commission, 2018.

² Medicare Payment Advisory Commission, 2019.

³ CMS, 2019a; CMS, 2019c; CMS, 2019d; CMS, 2019e; CMS, 2019f.

reimbursements to pay for patients in PAC settings;⁴ one goal of standardization is to allow for more alignment of payment models in the future.

Standardized data across PAC settings, facilitated by health information technology, could ensure the collection of high-quality, reliable information important for achieving person-centered outcomes and goals and improving care coordination. Interoperable, standardized data could also promote information exchange, enhancing efficiency through data sharing, and support decisionmaking. Providers could use standardized information to inform placements from acute care, coordinate transitions between PAC settings, and evaluate the performance of PAC settings for specific patient needs.

Project Overview

The Improving Medicare Post-Acute Care Transformation (IMPACT) Act, Section 2(a), mandates that CMS develop, implement, and maintain SPADEs for four PAC settings: HHAs, IRFs, LTCHs, and SNFs. Existing PAC assessment instruments by setting are the OASIS for HHAs, the IRF-PAI for IRFs, the LCDS for LTCHs, and the MDS for SNFs. SPADEs are to be nested within the four existing PAC assessment instruments; however, each instrument will continue to have distinct data elements selected for their special relevance to their respective PAC settings. The IMPACT Act mandates, at a minimum, SPADEs within the following clinical categories:

- functional status, such as mobility and self-care
- cognitive function and mental status
- special services, treatments, and interventions (SSTIs; e.g., need for ventilator, dialysis, chemotherapy, and total parenteral nutrition)
- medical conditions and comorbidities (e.g., diabetes, heart failure, and pressure ulcers)
- impairments (e.g., incontinence; impaired ability to hear, see, or swallow).

CMS contracted with the RAND Corporation to evaluate candidate SPADEs representing each of these categories (with the exception of functional status, which was addressed by another contractor), to meet the requirements of the IMPACT Act and because the SPADEs might support clinical decisionmaking, care coordination, cost reduction, and improved patient/resident and family experiences.

Stakeholder Input Opportunities

Development of standardized assessment data elements has required balancing the need to accommodate and adapt to the variations across PAC providers (e.g., in workflow) and the need to achieve a set of standardized data elements that can facilitate shared decisionmaking regardless of care setting. Maintaining an awareness of the diversity of environments (e.g.,

⁴ CMS, 2019b; CMS, 2019g; CMS, 2019h.

institution versus home setting) and patient/resident populations in which the assessments are implemented has been essential, as has the ongoing dialogue and support of the diverse body of stakeholders (patients, residents, family members, care providers, PAC administrators, association members).

Candidate SPADEs under each of the IMPACT Act categories were identified through an environmental scan, which included a literature review, consultation with experts in the field, input from the clinical communities serving the PAC populations (e.g., focus groups), discussions with stakeholders, discussions with partners within CMS and the U.S. Department of Health and Human Services, and feedback from our technical expert panel (TEP).

Prior to the National Beta Test, the Alpha 1 and Alpha 2 feasibility tests were conducted from August 2016 to October 2016 and from April 2017 to July 2017, respectively. Further information was collected from other stakeholder input opportunities, including two subregulatory calls for public comment and proposed rulemaking for the fiscal year (FY) 2018/calendar year (CY) 2019 rule cycle. See below for more detail on these activities and their corresponding reports:

- **TEP 1** (April 2016): Sixteen TEP members gathered for a two-day, in-person meeting to provide input on data elements in the current PAC assessments and other identified candidate SPADEs. TEP members rated data elements on their potential for improving quality, validity, feasibility for use in PAC, and utility for describing case mix. A report of the first TEP meeting can be found at https://www.rand.org/pubs/working_papers/WR1187.html.⁵
- **Public Comment 1** (PC1) (August to September 2016): A subregulatory call for public comment was solicited for a subset of the candidate SPADEs through the CMS website, for which 66 comments were received. The PC 1 summary report is available at <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/Development-of-Functional-Outcome-Quality-Measures-for-SNFs-Public-Comment-Summary-Report.pdf>.⁶
- **Alpha 1 Feasibility Testing** (August to October 2016): Alpha 1 was the first phase of pilot testing candidate SPADEs. Testing was conducted among four PAC providers (one of each PAC type) in the greater Hartford, Connecticut, area. Research nurses and facility/agency staff conducted 133 paired assessments so that results could be compared for both feasibility and interrater reliability. The Alpha 1 feasibility test report is available at <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/Standardized-Patient-Assessment-Data-Element-Alpha-1-Report.pdf>.⁷
- **TEP 2** (January 2017): Fourteen TEP members gathered for a two-day in-person meeting to review interim results of Alpha 1 testing and other potential candidate SPADEs. TEP

⁵ RAND Corporation, 2017a.

⁶ CMS, 2016.

⁷ Edelen et al., 2017.

members rated potential candidate SPADEs on their potential for improving quality, validity, feasibility for use in PAC, and utility for describing case mix. A follow-up webinar for TEP members was held to continue the discussion of candidate SPADEs (July 2017). A report of the second TEP meeting and follow-up webinar can be found at <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>.⁸

- **Public Comment 2 (PC2)** (April to June 2017): A second subregulatory call for public comment was solicited through the CMS website, for which 33 comments were received. The PC 2 summary report is available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/Public-Comment-Summary-Report_Standardized-Patient-Assessment-Data-Element-Work_PC2.pdf.⁹
- **Proposed rulemaking** (fall 2016 to summer 2017): Some candidate SPADEs that had previously undergone feasibility testing (i.e., before the Alpha 1 pilot commenced) were submitted for proposed rulemaking for the FY 2018/CY 2019 rule cycle. Results of this process are available in the *Federal Register* at <https://www.federalregister.gov/d/2017-07800> (for LTCHs); <https://www.federalregister.gov/d/2017-08521> (for SNFs); <https://www.federalregister.gov/d/2017-08428> (for IRFs); and <https://www.federalregister.gov/d/2017-15825> (for HHAs).¹⁰
- **Alpha 2 Feasibility Testing** (April to July 2017): Testing was conducted among 15 PAC providers in three regions of the United States. As with Alpha 1 testing, research nurses and facility/agency staff conducted paired assessment (for communicative patients, 118 at admission and 42 at discharge; for non-communicative patients, 44 assessments) so that results could be compared for both feasibility and interrater reliability. The Alpha 2 feasibility test report is available at <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/Alpha-2-SPADE-Pilot-Summary-Document.pdf>.¹¹

The results of these activities combined to inform the content and design of the National Beta Test to evaluate candidate SPADE performance when used in any of the four PAC settings. The overarching goal of the National Best Test is to evaluate the reliability and validity of candidate data elements and identify the best, most feasible subset for standardization to meet requirements of the IMPACT Act.

Stakeholder engagement activities continued throughout the National Beta Test field period and beyond:

⁸ RAND Corporation, 2017b.

⁹ CMS, 2018.

¹⁰ CMS, 2017a, 2017b, 2017c, 2017d.

¹¹ Edelen et al., 2018a.

- **CMS/RAND stakeholder forum** (November 2018): CMS and RAND hosted a public forum to present an overview of the results from the National Beta Test and answer questions from in-person and virtual attendees. Presentation materials, a transcript of the session, and frequently asked questions related to the presentation are available on CMS’s IMPACT Act Archived Information webpage:
 - Forum slide deck: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/Forum-Slides_Early-Findings-from-the-RAND-IMPACT-National-Beta-Test.pdf¹²
 - Forum transcript: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/TTranscript-of-11_27_18-Forum.pdf¹³
 - Forum frequently asked questions: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/QA_RAND-Forum-on-PAC-SPADES.pdf¹⁴
 - Following the forum, CMS invited written, emailed feedback on the presentation. A verbatim comment summary of this feedback is also available online: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/Input-on-SPADEs-Received-After-November-27-2018-Stakeholder-Meeting.pdf>.¹⁵

Candidate SPADEs in the National Beta Test

The National Beta Test was built on early information-gathering activities in the project and the Alpha 1 and Alpha 2 pilot tests. This section describes the candidate SPADEs included in the National Beta Test and the activities through which quantitative and qualitative feedback were gathered on them. Candidate SPADEs in the National Beta Test were identified in consultation with CMS and after a rigorous review and development process. For some candidate SPADEs, this process included refinement in response to feedback obtained through stakeholder engagement and testing.

Cognitive Function and Mental Status

Cognitive Function

Conducting cognitive assessments in PAC settings is essential to screen for cognitive impairment, rate severity of disorder, and develop a plan for care transitions. Therefore, the following candidate SPADEs that assess cognitive status were included in the National Beta Test:

¹² RAND Health Care, 2018.

¹³ Edelen et al., 2018b.

¹⁴ RAND Corporation, 2018.

¹⁵ RAND Corporation, 2019.

- Brief Interview for Mental Status (BIMS)
- Confusion Assessment Method (CAM)
- Staff Assessment of Mental Status
- Expression and Understanding: (1) Speech Clarity, Makes Self Understood, and Ability to Understand Others; (2) Expression of Ideas and Wants, and Understanding Verbal Content
- Behavioral Signs and Symptoms (Presence and Frequency, Impact on Patient/Resident, Impact on Others, Rejection of Care) to assess whether the patient/resident has exhibited any behavioral symptoms that may indicate cognitive impairment or other issues during the assessment period.

Mental Status

Mental health issues are associated with poor outcomes and slowed recovery among older adults in PAC settings. Four candidate SPADEs that assess mental status were included in National Beta Testing:

- Patient/Resident Mood Interview (Patient Health Questionnaire [PHQ]-2 to 9)
- Staff Assessment of Patient/Resident Mood (PHQ-9-OV)
- Patient-Reported Outcomes Measurement Information System (PROMIS) Anxiety
- PROMIS Depression.

Special Services, Treatments, and Interventions

Assessing patients/residents for the use of SSTIs in PAC settings provides important information about the severity of illness, the risk of complications and adverse health outcomes, and the intensity of resource use.

Nutritional Approaches

Four candidate SPADEs for assessment of Nutritional Approaches were included in the National Beta Test:

- Parenteral/Intravenous (IV) Feeding
- Feeding Tube
- Mechanically Altered Diet
- Therapeutic Diet.

Special Treatments

The National Beta Test included several candidate SPADEs for assessment of Special Treatments:

- Chemotherapy
- Radiation
- Oxygen Therapy
- Suctioning
- Tracheostomy Care

- Invasive Mechanical Ventilator
- Non-Invasive Mechanical Ventilator
- IV Medications
- Transfusions
- Dialysis
- IV Access.

Medical Conditions—Pain

Pain significantly adversely affects a person’s quality of life and is linked to poor mental and physical health outcomes. The following interview and observational candidate SPADEs to assess pain were included in the National Beta Test:

- Pain Interview (including pain presence, frequency, severity, impact on sleep, therapy and nontherapy activities, and relief from pain due to treatment)
- Observational Assessment of Pain or Distress.

Impairments

Hearing and Vision

Hearing and vision impairments are common conditions among older adults. If unaddressed, these impairments can create difficulty communicating and impede accurate assessment of other medical conditions, such as cognition. The following candidate SPADEs that assess hearing and vision were included in the National Beta Test:

- Ability to Hear
- Ability to See in Adequate Light.

Bladder and Bowel Continence

Impaired bowel and bladder continence are common conditions that, if unaddressed, can affect a patient’s/resident’s activities of daily living, rehabilitation outcomes, skin integrity, or overall quality of life. The National Beta Test included the following candidate SPADEs:

- Experience and Perceived Problem or Burden with Bladder Incontinent Events
- Experience and Perceived Problem or Burden with Bowel Incontinent Events
- Bladder Appliance Use
- Bowel Appliance Use
- Bladder Frequency of Incontinent Events
- Bowel Frequency of Incontinent Events.

Other Clinical Categories

Global Health

Patient-reported outcomes include direct reports from patients/residents about their function, symptoms, and perceptions of their health and/or response to therapies. The goal of health-

related quality of life assessment is to quantify, in a valid and reproducible way, the degree to which a medical condition or its treatment affects an individual's life.

The National Beta Test included the following candidate SPADE:

- PROMIS Global Health.

Care Preferences

The assessment of patient/resident care preferences and goals for care is critical to ensuring patient-centered and preference-concordant care through the course of a PAC episode and beyond. The following candidate care preferences SPADEs were included in the National Beta Test:

- Importance of involvement of family/friends (interview)
- Preferences for involvement in decisionmaking (interview)
- Advance directive—health care agent (chart review).

Medication Reconciliation

Medication reconciliation, the process of obtaining a patient's/resident's multiple medication lists and reconciling any discrepancies, is an effective way to promote patient/resident safety by reducing errors and resulting adverse drug events. Medication reconciliation candidate SPADEs included in the National Beta Test use multiple information sources to assess

- the classes of medications a patient/resident is currently taking
- whether an indication is noted for medications
- whether there were discrepancies
- whether a patient/resident or family/caregiver was involved in addressing discrepancies
- whether discrepancies were communicated to a physician within 24 hours
- whether recommended physician actions regarding discrepancies were carried out within 24 hours
- whether the reconciled medication list was communicated to a patient/resident, prescriber, and/or pharmacy.

2. Methods

The National Beta Test was designed to collect data from patients and residents receiving care across four PAC settings (HHAs, IRFs, LTCHs, and SNFs), allowing for thorough evaluation of candidate SPADEs. Developing SPADEs suitable for cross-setting use requires establishing the reliability and validity of the items and scales. It was important that we understand the performance of data element properties both in ideal circumstances and in the types of conditions that are likely to be present in real-life practice or situations. Thus, to be successful, the field test design needed to include clear instructions and data collection protocols, the ability to test performance of “gold-standard” nurse assessors and actual staff across care settings, and the ability for robust evaluation of the performance of data elements across the diverse patient/resident populations served by SNFs, HHAs, IRFs, and LTCHs. In the following section, we briefly review the many aspects of the methods used in the National Beta Test data collection and evaluation. More-detailed information on each of these aspects is available in Volume 2.¹⁶

Design of Data Collection

Assessment data were collected by trained research nurses in each region, or “market,” and trained facility/agency staff from each participating provider. The role of the research nurse was to oversee field data collection in their market and to serve as a “gold standard” assessor alongside a facility/agency staff person, allowing measurement of interrater reliability.

Testing included three types of protocols with slightly different focuses to test the performance of candidate SPADEs for different types of patients/residents at both admission and discharge.¹⁷ The first protocol, the communicative admission assessment, and the second protocol, the communicative discharge assessment, collected data from patients/residents who could make themselves understood using any means (i.e., writing, gesturing, speaking) at admission and discharge, respectively. The third protocol, the non-communicative assessment, was administered at any point during a qualifying PAC stay for non-communicative patients/residents who were unable to make themselves understood in any fashion. Trained research nurses and/or staff at participating PAC facilities and agencies administered all National Beta Test assessment protocols. A subset of National Beta Test assessments was completed by research nurse and facility or agency staff assessor pairs to allow for evaluation of interrater reliability. Other National Beta Test design features allowed for comparison of different look-

¹⁶ Edelen et al., 2019a.

¹⁷ CMS, 2017c.

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. We set data collection target completion numbers for each type of assessment and by PAC provider type, based on the number we estimated would be needed to support the planned data analysis.

Market and Facility Sampling and Recruitment

A multi-stage stratified random sampling plan was used to obtain the sample of 14 geographic/metropolitan hospital referral regions, or “markets,” in the United States, and then a sample of eligible PAC facilities (e.g., based on size and location) was compiled from those markets. The sampling plan is described in more detail in Volume 2. National Beta Test data were collected from participating PAC facilities/agencies in the following 14 markets: Boston, Massachusetts; Chicago, Illinois; Dallas, Texas; Durham, North Carolina; Fort Lauderdale, Florida; Harrisburg, Pennsylvania; Houston, Texas; Kansas City, Missouri; Los Angeles, California; Nashville, Tennessee; Philadelphia, Pennsylvania; Phoenix, Arizona; St. Louis, Missouri; and San Diego, California.

Within each market, all providers of each PAC type that met eligibility criteria (e.g., threshold number of Medicare patients; see Volume 2) were identified from Medicare administrative files, and a subset were sampled for recruitment. If attempts to recruit sufficient providers from the initial sample were exhausted, additional facilities/agencies were sampled for recruitment on an as-needed basis throughout the recruitment phase to meet the target numbers of settings in each market. Provider recruitment began in March 2017 and ended in December 2017 and consisted of an in-person informational event, small group outreach, informational packets sent by mail, and follow-up telephone calls. These efforts resulted in 186 providers agreeing to participate in the National Beta Test and participating in training (57 HHAs, 28 IRFs, 28 LTCHs, and 73 SNFs).

Assessor Training

The National Beta Test design planned for each of the 14 markets to be supported by two research nurses who reside in the area (for a total of 28). Research nurses were assigned to sites in their market and worked in partnership with the facility/agency staff at their sites to collect data. Research nurse recruitment occurred between June and September of 2017. Research nurses were successfully recruited for all markets prior to the start of research nurse training. Replacement hiring occurred as needed throughout the project period.

To ensure rigorous data collection during the National Beta Test, we implemented a comprehensive multi-component training plan, consisting of virtual and in-person research nurse training and 18 market-specific (i.e., local) in-person facility/agency staff trainings. Trainings for all research nurses and approximately 350 facility/agency staff nationwide began in September

2017 and concluded in March 2018. Each component of the National Beta Test training plan is described below.

Research nurse training consisted of web-based e-modules to be completed prior to attending the in-person component; a pre-training webinar; a week-long in-person training on the administration of the candidate SPADEs, including hands-on practice with the assessment tool and electronic data collection procedures; and a webinar on repeat assessment data collection.

Facility/agency staff were required to attend an in-person training held over two days at a central location in their market prior to beginning assessments. In-person training was conducted by project staff from RAND and Qualidigm, with assistance from the recently trained research nurses who would be working in the markets. Following the in-person training, facility/agency staff were required to conduct at least two practice assessments at their site to augment their training within the particular context of their site, with guidance and oversight from their research nurse partner.

Refresher and replacement trainings were held later in the data collection period as needed because of the potential time lag between training and data collection and to accommodate turnover in research nurses and field staff.

Data Collection

Data collection began after training and practice assessments were completed in each market, between November 2017 and February 2018, and concluded on August 15, 2018. Within each market, the research nurses who were hired to work in those markets were assigned four to seven PAC providers each, with whom they worked to collect assessment data. Research nurse assignment was done by clustering facilities/agencies into geographic areas within markets to minimize travel time for the research nurses. Data collection was done electronically using encrypted, handheld tablets.

Research nurses and facility/agency staff were given support throughout the data collection period in the form of a market manager, who tracked weekly data collection progress; a regional manager, who helped with troubleshooting and supported networking among the research nurses; and weekly market-specific calls to review progress and challenges for each site. RAND also maintained a project website with training and data collection materials and promptly responded to phone and email contact from field staff to resolve issues.

Given the lengthy data collection period, priority was placed on ensuring ongoing engagement and retention of National Beta Test providers. To this end, several activities were implemented, including a monthly project newsletter, provider webinars, and outreach to corporate leadership. All staff followed a strict data safeguarding protocol to protect exposure of personal information.

Completed assessments were uploaded to a secure server by field staff. RAND Survey Research Group staff downloaded the assessments and completed a quality assurance check. The

Survey Research Group periodically delivered a cumulative file of all submitted assessments received to the analytic team, who conducted additional data cleaning and validation activities.

In addition to collecting and analyzing quantitative assessment data during the National Beta Test, we also collected qualitative feedback from facility/agency staff and research nurses to augment our understanding of workflows, barriers, and challenges to implementing the assessment in practice and burden on staff. We administered surveys to facility/agency staff and conducted separate focus groups with both field staff and research nurses to collect this information.

Analytic Plan

There were four primary goals for the data analysis: (1) to compute descriptive statistics for candidate SPADEs, (2) to determine the feasibility of administration, (3) to evaluate interrater reliability, and (4) to determine the content validity of candidate SPADEs. If candidate SPADEs were included in the supplemental components of the National Beta Test (e.g., Day 3, 5, 7 test), additional analysis goals included (5) to understand the best look-back time frame and assessment days, (6) to identify the most appropriate versions of candidate SPADEs for cross-setting standardization, and (7) to understand the psychometric properties of candidate SPADEs that are possible components of multi-data element scales, such as from the PROMIS item banks. We took the following steps to accomplish these goals:

1. Response frequencies were computed for each candidate SPADE using data collected from facility/agency staff and unpaired research nurse admission assessments to provide descriptive information. Where relevant, statistical comparisons were reported on tabulated data for prespecified groupings (e.g., provider type, gender, clinical conditions).
2. Feasibility of administration was informed by the extent of missing data and the time to complete. Completion times were estimated based on the subset of assessments completed by facility/agency staff, because they are the primary group of interest for evaluating the ease of administration. Time spent to complete the items was self-reported by facility/agency staff and collected via time stamps recorded on tablets during the assessment administration. Facility/agency staff reports were favored over tablet time stamps, when available.
3. To determine whether data elements could be completed with acceptable interrater reliability, we calculated the level of agreement between pairs of research nurse and facility/agency staff assessors. For dichotomous data, we computed kappa and overall agreement percentage. For ordinal data element data, we computed weighted kappa. For continuous or approximately continuous data (more than five ordered categories), we computed inter-class correlation coefficients. Additionally, because kappa is sensitive to base or prevalence rates, we report raw percent agreement as a supplemental/alternative index of interrater reliability for all data elements.
4. To consider validity of candidate SPADEs, we examined associations between National Beta Test data elements that should be related based on literature and clinical practice (convergent validity) and the extent to which a data element represents all facets of a

given construct consistently across a variety of relevant populations (content or known groups validity).

5. We employed a variety of approaches to evaluate the performance of different look-back time frames and assessment days.
 - a. For chart review SPADEs that were assessed over different look-back time frames (Day 1, 3, 5, 7), we evaluated cumulative percentages to observe rates of increase. Then, we used cumulative percentages to compute the ratio of percent occurrence for each Day (1, 3, 5) to the total percent occurrence by Day 7. The resulting values were used to determine whether rates of occurrence depend on the look-back period and, if so, to identify the earliest chart review day that captures the majority of occurrences.
 - b. For candidate SPADEs included in repeat assessments, the goal was to understand whether it matters which day patients/residents are assessed (e.g., at Day 3, Day 5, or Day 7). Thus, analyses examined whether there are significant and meaningful differences in rates or scores depending on the day a patient/resident is assessed and whether that varies by setting.
 - c. We conducted similar evaluations to determine the extent of stability or change over time from admission to discharge for all candidate SPADEs, except those developed superficially for use with non-communicative patients/residents.
6. In cases where we tested alternate forms of data elements, we attempted to determine which form (version) of the candidate SPADE may be most appropriate for cross-setting standardization by examining the distribution of scores on these data elements and overall scale scores (where relevant), as well as interrater reliability, to make comparisons between alternate forms.

To understand the psychometric properties of scalable data elements (e.g., PROMIS Anxiety), we examined basic data element–level and scale-level properties, such as internal consistency reliability, unidimensionality, and differential item functioning.

Sample

A total of 143 PAC facilities/agencies (35 HHAs, 22 IRFs, 26 LTCHs, and 60 SNFs) contributed patient/resident assessment data to the National Beta Test. Patients and residents who were receiving care at one of these 143 participating provider sites and were Medicare beneficiaries covered under one of the PAC prospective payment systems were eligible for inclusion. In all, the National Beta Test included 3,669 patients/residents, 3,121 of whom comprised the communicative admission sample and 548 of whom comprised the non-communicative sample. Total counts for each assessment type submitted overall and by setting are shown in Table 2.1, along with the average number of submitted assessments overall and per setting type.

Table 2.1. Total and Average Assessment Counts by National Beta Test Subsample, Overall and by PAC Setting Type

Assessment Type	HHA	IRF	LTCH	SNF	Overall
Communicative admission	653 (19)	794 (36)	507 (20)	1,167 (20)	3,121 (22)
Interrater reliability	199 (6)	261 (12)	242 (9)	274 (5)	976 (7)
Day 3, 5, 7 repeat assessment	112 (3)	150 (7)	91 (4)	239 (4)	592 (4)
Discharge	148 (4)	350 (16)	90 (4)	235 (4)	823 (6)
Non-communicative	32 (1)	107 (5)	185 (7)	224 (4)	548 (4)
Total patients/residents assessed ^a	685 (20)	901 (41)	692 (27)	1,391 (23)	3,669 (26)

NOTES: Averages are given in parentheses. Interrater reliability; Day 3, 5, 7 repeat assessments; and discharge assessments were all conducted among subsets of patients/residents with a communicative admission assessment.

^a This total row is the sum of the communicative admission and non-communicative rows.

The setting totals were in line with recruitment targets (e.g., the National Beta Test included relatively more SNFs and HHAs but still included more than 20 IRFs and LTCHs), and although there was some variability in the number and type of participating facilities/agencies according to market, much of this variability reflected true variability in the population (e.g., nationally there are proportionally more LTCHs in Texas than in other markets, so the Dallas and Houston markets in the National Beta Test had proportionally more LTCHs than other setting types), and in general there was fairly even spread across the markets, with half of the 14 markets having at least one of each setting, and only one market (Nashville) contributing facilities/agencies across only two setting types (HHA and SNF). Assessment counts were fairly evenly distributed according to data collection targets, with the exceptions that IRFs, on average, tended to contribute more communicative assessments than the other three settings and that LTCHs tended to contribute more non-communicative assessments.

The population of facilities and agencies in the 14 sampled markets is fairly representative of the national population of PAC facilities/agencies. However, because of study design constraints, the National Beta Test sample had very low rural representation, and larger facilities/agencies were somewhat overrepresented.

Despite this imbalance in size and urbanicity of participating facilities/agencies, the National Beta Test patient/resident sample was comparable to the national population on patient/resident demographics and clinical conditions. However, the National Beta Test patient/resident sample had slightly higher rates of stroke, sepsis, and heart failure than the national population of PAC patients/residents.

Finally, with regard to the representativeness of the National Beta Test communicative subsamples, although there were some significant differences in characteristics according to

subsample inclusion, these differences tended to be limited and relatively small. This general comparability lends strength to the results and conclusions regarding candidate SPADE performance in this National Beta Test.

3. Results

In this chapter, we first review the basic feasibility and reliability results for each of the candidate SPADEs representing each clinical category evaluated in the National Beta Test. SPADEs developed specifically for patients/residents who are unable to communicate are grouped together and reviewed last. This is followed by a summary of results across SPADEs pertaining to evaluation of validity, look-back differences (Day 3, 5, 7 repeat assessment and Day 1, 3, 5, 7 chart review), and stability/change from admission to discharge.

Results for Data Elements Related to Cognitive Function

The BIMS, CAM, and Expression and Understanding data elements were selected for testing in the National Beta Test based on information-gathering activities, expert and stakeholder input, and Alpha feasibility testing. Results of their performance are reviewed below.

Brief Interview for Mental Status

The BIMS is a performance-based cognitive assessment that assesses repetition, recall with and without prompting, and temporal orientation. The BIMS is completed via patient/resident interview to determine how the patient/resident performs on a series of tasks. The BIMS is currently used in the MDS and IRF-PAI, and an image of the BIMS as tested in the National Beta Test can be found in Volume 4 (www.rand.org/t/RR3004z3).¹⁸ In the National Beta Test, the BIMS was included in the repeat assessment evaluation and therefore was evaluated repeatedly on the same patient/resident by the same assessor on Days 3, 5, and 7, per the repeat assessment design.

Results

Assessing impairment: In the National Beta Test, the BIMS was administered at admission to 3,062 patients/residents across the four settings. Five percent of patients/residents met criteria for being severely impaired, 18 percent moderately impaired, and 76 percent intact. SNFs and LTCHs had more cognitively impaired patients/residents compared with IRFs and HHAs.

Missing data: Overall, data element–level missing data ranged from 0.4 to 1.7 percent, with minimal setting differences. For all settings, missing data rates were slightly higher for recall of current day of the week than for other data elements.

¹⁸ Edelen et al., 2019b.

Time to complete: Overall mean time to complete the BIMS was 2.2 minutes, with some setting variability. Specifically, it took significantly less time to complete the BIMS in IRFs (1.8 minutes) than in any other setting.

Known groups validity: We hypothesized that cognitive impairment would be associated with age, heart failure, stroke, needing assistance with toileting, and ability to transfer from lying to sitting. Analyses of these associations using BIMS impairment categorization (intact, moderately impaired, severely impaired) supported four of these five hypotheses.

Interrater reliability: The interrater reliability was excellent for the BIMS as measured by kappa (0.91 for the BIMS Impairment Category classification, and 0.83 to 0.93 for individual data elements within the BIMS) and percent agreement (96 percent for the Impairment CATEGORY, and 94 to 98 percent for individual data elements) of paired raters. In each of the four settings, kappas were excellent (greater than or equal to 0.81), except for one data element in SNFs (0.78: “recalls blue”). Similarly, percent agreement was excellent in each of the four settings, ranging from 91 to 99 percent.

Assessor feedback: National Beta Test facility/agency staff assessors rated the BIMS among the top five data elements in terms of clinical utility and as one of the least burdensome data elements for both assessors and patients/residents. In focus groups, respondents noted the brevity and widespread use of the BIMS, suggesting that it might be a good candidate for standardization across PAC provider types. Although both facility/agency staff and research nurses said that frequent administration of the BIMS had caused some patients/residents to memorize the recall component (“bed, sock, and blue”), it was further noted that this frequency of assessment was an artifact of the National Beta Test design and would not pose a significant problem in practice.

Confusion Assessment Method

The CAM screens for overall cognitive impairment and includes features to distinguish delirium or reversible confusion from other types of cognitive impairment.¹⁹ Specifically, the CAM screens for acute change in mental status, inattention, disorganized thinking, and altered level of consciousness and, if exhibited, documents whether these behaviors are continuously present or fluctuating. Versions of the CAM are currently in the MDS and LCDS, and an image of the CAM as tested in the National Beta Test can be found in Volume 4 (www.rand.org/t/RR3004z3).

Results

Assessing delirium: The CAM was administered at admission to 2,973 patients/residents across the four settings. Five percent of patients/residents had evidence of mental status change from baseline, 12 percent had difficulty focusing (3 percent continuously), 6 percent had

¹⁹ The CAM is under separate copyright protection. © 1988, 2003, Hospital Elder Life Program. All rights reserved. Adapted from Inouye et al., 1990. The Hospital Elder Life Program, LLC, has granted permission to use the CAM in association with the PAC instruments.

disorganized thinking (1 percent continuously), and 4 percent had altered consciousness (1 percent continuously). Rates were generally similar across settings; however, compared with other settings, difficulty focusing was slightly higher in IRFs (15 percent) and altered consciousness was slightly lower in HHAs (2 percent).

Missing data: Overall, there were very low rates of missing data for the CAM, not higher than 0.4 percent for any data element and with limited variability across settings.

Time to complete: Overall mean time to complete the CAM was 1.4 minutes across settings, with some setting variability. Specifically, it took significantly less time to complete the CAM in IRFs (1.3 minutes) than in any other setting.

Known groups validity: We hypothesized that a change in mental status as assessed by the CAM would be associated with age and stroke. Analyses of these associations supported one of these two hypotheses.

Interrater reliability: The interrater reliability was good for the CAM, as measured by kappa and percent agreement of paired raters. The kappa for the focusing attention data element was good (0.66) overall, but kappas were not able to be estimated for the other three data elements given that prevalence rates were out of range for stable kappa estimates as determined by study power calculations. Similarly, kappa for the focusing attention data element was good in all settings (0.66 to 0.75) except in IRFs, where it was moderate (0.55). Additionally, kappas were good for the change in mental status data element in IRFs (0.60) and disorganized thinking data element in SNFs (0.60), but kappas were not able to be estimated in the other settings given prevalence rates. Percent agreement for the CAM across settings was high for all four CAM data elements (91–96 percent). Similarly, percent agreement was high in each of the four settings (89–98 percent).

Assessor feedback: According to the assessor survey, facility/agency staff and research nurses found the CAM to be moderately clinically useful, although there was some variation in opinion about the overall usefulness of the CAM for cross-setting use. Specifically, facility assessors in LTCHs rated the CAM's clinical utility higher than other settings. Both facility/agency staff and research nurses considered it to be a relatively low-burden data element.

Expression and Understanding

The Expression and Understanding data elements assess whether a patient/resident is able to express or communicate requests, needs, and opinions; conduct social conversation; and comprehend direct person-to-person communication. Currently, two data elements, Expression of Ideas and Wants and Understanding Verbal Content, are included in the IRF-PAI and LCDS. The MDS includes similar data elements—Makes Self Understood and Ability to Understand Others—and also includes Speech Clarity. In the National Beta Test, we tested two versions of data elements, corresponding to the forms in current use in the MDS (version 1, three data elements), and IRF-PAI and LCDS (version 2, two data elements). Images of both versions as tested in the National Beta Test can be found in Volume 4 (www.rand.org/t/RR3004z3).

Results

Assessing expression and understanding: The three–data element version of the Expression and Understanding data elements was administered to 1,534 patients/residents. The two–data element version was administered to 1,529 patients/residents. For the three– and two–data element versions, 90 percent of patients/residents were “understood” or “expressed without difficulty,” and 88 percent and 89 percent of patients/residents “understood verbal content” or “understood without cues or repetitions,” respectively. In HHAs, however, fewer patients/residents who received the three–data element version were “understood” (74 percent) or “understood verbal content” (69 percent) compared with patients/residents who received the two–data element version (90 and 86 percent, respectively). For patients/residents who received the three–data element version, 95 percent exhibited clear speech.

Missing data: Missing data for Expression and Understanding data elements ranged from 0.0 to 0.1 percent across the two versions, with minimal setting differences.

Time to complete: On average, the three–data element version of Expression and Understanding took 0.8 minutes to complete, with only LTCHs taking significantly more time to complete (0.9 minutes) than HHAs (0.8 minutes). The two–data element version took 0.7 minutes to complete, with only HHAs taking significantly more time to complete (0.8 minutes) than IRFs (0.7 minutes).

Known groups validity: We hypothesized that the Expression and Understanding data elements would be associated with stroke, toileting, and ability to transfer from lying to sitting. Analyses of these associations supported all three of these hypotheses for both the Expression data elements and the Understanding data elements.

Interrater reliability: For the three–data element version overall, kappa was good (0.64) for “ability to express ideas and wants” and moderate (0.59) for “understanding verbal content.” At the setting level, kappa was moderate (0.50) for “ability to express ideas and wants” in HHAs and moderate for “understanding verbal content” in all settings except IRFs (0.43 to 0.60). Remaining setting-level kappas were not estimated, given that prevalence rates were out of range for stable kappa estimates as determined by study power calculations. For the two–data element version overall, kappa was moderate (0.42) for “expresses ideas and wants” and fair (0.32) for “understands verbal content.” At the setting level, kappa for “expresses ideas and wants” was fair to moderate (0.26 to 0.58) and fair for “understands verbal content” (0.22 to 0.39). Overall percent agreement for the three–data element version was high for all Expression and Understanding data elements, ranging from 93 to 95 percent, while overall percent agreement was slightly lower for the two–data element version, with 89 percent and 86 percent for expression and understanding, respectively. At the setting level, percent agreement for the three–data element version ranged from 81 to 97 percent (slightly lower in HHAs), and percent agreement for the two–data element version ranged from 82 to 93 percent (slightly lower in LTCHs).

Assessor feedback: Assessors noted the clinical utility of having consistent information about patients'/residents' ability to express themselves and be understood; in the assessor survey, these data elements were rated among the top five data elements in terms of clinical utility regardless of which version was used. In addition, both facility/agency staff and research nurses considered these data elements to have low burden.

Behavioral Signs and Symptoms

The Behavioral Signs and Symptoms data elements assess the presence and frequency of behavioral symptoms and the impact of behavioral symptoms on the patient/resident and on others by reviewing the medical record, interviewing staff and others who interact with the patient/resident, and observing the patient/resident. Behavioral symptoms are currently assessed by the OASIS and MDS; the version of the data elements tested in the National Beta Test was derived from the MDS, and an image of the data elements as tested in the National Beta Test can be found in Volume 4 (www.rand.org/t/RR3004z3).

Results

Assessing behavioral signs and symptoms: Frequency data for the Behavioral Signs and Symptoms data elements were collected on 2,954 patients/residents. Rejection of evaluation and care, the most frequently observed behavior, was exhibited by 5.5 percent of all patients/residents and ranged from 3.8 percent in HHAs to 8.3 percent in SNFs. Overall, the other behavioral signs and symptoms were exhibited by less than 2 percent of patients/residents, with consistently lower rates in HHAs and SNFs.

Missing data. Missing data for the Behavioral Signs and Symptoms data elements ranged from 0.2 to 0.5 percent, with minimal setting differences.

Time to complete: Overall mean time to complete the Behavioral Signs and Symptoms data elements was 1.4 minutes, with minimal setting differences.

Known groups validity: We hypothesized that Behavioral Signs and Symptoms would be associated with length of stay, toileting, and ability to transfer from lying to sitting. Analyses of these associations using any behavioral symptom exhibited (physical, verbal, or other) did not support any of these three hypotheses.

Interrater reliability. Kappa interrater reliability coefficients are not shown for the Behavioral Signs and Symptoms data elements overall or by setting, given that prevalence rates were out of range for stable kappa estimates as determined by study power calculations. Overall percent agreement was high for all Behavioral Signs and Symptoms data elements, ranging from 95 to 100 percent. Similarly, percent agreement was excellent in each of the four settings, ranging from 93 to 100 percent.

Assessor feedback: Assessors mentioned that information about behaviors is clinically relevant and valuable for effective transfers across PAC settings. They noted that standardization would increase continuity of care and better prepare the staff at the next site of care. The

SPADEs were perceived as having moderately low burden. However, some assessors noted that documentation of behavioral signs and symptoms can be inconsistent because a formal record of disruptive behaviors may prevent a transfer to another PAC provider.

Results for Data Elements Related to Mental Status

The PHQ-2 to 9, PROMIS Depression, and PROMIS Anxiety data elements were selected for testing in the National Beta Test based on information-gathering activities, expert and stakeholder input, and Alpha feasibility testing. Results of their performance are reviewed below.

PHQ-2 to 9

The PHQ-2 to 9 uses the first two elements (PHQ-2) as a gateway for the longer PHQ-9; the assessor administers the full PHQ-9 only if the initial score on the PHQ-2 exceeds a threshold indicating possible depression.²⁰ An image of the PHQ-2 as tested in the National Beta Test can be found in Volume 5 (www.rand.org/t/RR3004z4).²¹

Results

Assessing depressed mood: In National Beta Testing, the PHQ-2 to 9 was administered to 3,010 patients/residents across settings. Thirty-eight percent of patients/residents reported having little interest in doing things, and 43 percent reported feeling down, depressed, or hopeless at some point in the past 14 days. More than one in four patients/residents (28 percent) across settings screened positive on the PHQ-2 and completed all nine data elements. This rate was higher in LTCHs (38 percent) than in other settings (24 to 27 percent). The average PHQ-2 score across settings for those who screened negative was 2.4. For patients/residents who completed the full PHQ-9, the average score across settings was 11.9. In LTCHs, the average full PHQ-9 score (13.0) was significantly higher than in SNFs (11.5).

Missing data: Across all settings, data element-level missing data for the PHQ-2 to 9 did not exceed 5.2 percent for any of the data elements and varied minimally across settings.

Time to complete: Among patients/residents who only received the PHQ-2, time to complete was an average of 1.7 minutes overall and took significantly less time to complete in IRFs (1.5 minutes) than in LTCHs (1.9 minutes) and HHAs (1.8 minutes). Among those receiving the full PHQ-9, time to complete was an average of 4.0 minutes, with minimal settings differences.

Known groups validity: We hypothesized that signs and symptoms of depression as indicated by patient/resident response to the PHQ-2 to 9 would be associated with gender, age, stroke, toileting, and ability to transfer from lying to sitting. Analyses of these associations supported four of these five hypotheses for either a positive screen on the PHQ-2 or total PHQ-9 scores.

²⁰ Spitzer, Kroenke, and Williams, 1999. The Patient Health Questionnaire (PHQ) was developed by Pfizer Inc. © 1999 Pfizer Inc. All rights reserved.

²¹ Edelen et al., 2019c.

Interrater reliability: Interrater reliabilities for all symptom presence and frequency data elements, the PHQ-2, and the full PHQ-9 were nearly perfect: kappas ranged from 0.95 to 1.00 for the four settings combined. Similarly, in each of the four settings, kappas were excellent (greater than or equal to 0.90). Percent agreement was also nearly perfect for all symptom presence and frequency data elements, the PHQ-2, and the full PHQ-9, ranging from 95 percent to 100 percent overall. Similarly, percent agreement was excellent in each of the four settings, ranging from 93 to 100 percent

Assessor feedback: Assessors thought the PHQ-2 to 9 was clinically relevant; assessors rated the PHQ-2 to 9 in the middle to high range on clinical utility in the assessor survey, relative to other data elements. Although there was a general appreciation for the range of questions asked in the PHQ-2 to 9, facility/agency staff focus group participants felt that neither PHQ-2 to 9 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). Assessors noted that the two-week recollection period was difficult for some patients/residents, and it was possible that this period included an acute care stay.

PROMIS Depression

The PROMIS Depression data elements fielded in the National Beta Test included eight data elements selected from among the 28 data elements that make up the PROMIS Depression Item Bank, which assesses depression across a wide range of symptoms.²² These data elements are not currently used in PAC patient/resident assessment instruments. In the National Beta Test, the PROMIS Depression data elements were collected in two versions, which used a reference to “the past 7 days” or “the past 3 days” for symptoms. Images of both versions as tested in the National Beta Test can be found in Volume 5 (www.rand.org/t/RR3004z4).

Results

Assessing depressed mood: The “past 7 days” version of the PROMIS Depression data elements was administered to 1,498 patients/residents. The “past 3 days” version was administered to 1,488 patients/residents. Overall, rates were similar between versions and tended to be highest for data elements asking about feeling sad (63 and 61 percent for the “past 7 days” and “past 3 days” versions, respectively) and helplessness (55 and 52 percent, respectively) and tended to be lowest for data elements asking about no reason for living (17 and 19 percent, respectively) and hopelessness (30 and 29 percent, respectively). Overall depression total scores were also similar for both versions (14.7 and 14.8 percent, respectively). For all data elements and depression total score, LTCHs consistently had higher rates and scores (on both versions) compared with all other settings.

²² 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.

Missing data: Overall missing data for the PROMIS Depression data elements ranged from 0.8 percent to 2.14 percent in the “past 7 days” version and 0.81 percent to 1.48 percent in the “past 3 days” version, with minimal setting differences.

Time to complete: On average, the time to complete the PROMIS Depression data elements for the “past 7 days” version and the “past 3 days” version was the same at 2.2 minutes, with some variability across settings. For the “past 3 days” version, IRFs took significantly less time to complete (1.9 minutes) the PROMIS Depression data elements compared with HHAs (2.4 minutes) and SNFs (2.4 minutes).

Known groups validity: We hypothesized that depression would be associated with gender, age, stroke, toileting, and ability to transfer from lying to sitting. Analyses of these associations using the PROMIS Depression t-score supported four of these five hypotheses.

Interrater reliability: Kappas for the PROMIS Depression data elements were excellent overall, ranging from 0.97 to 0.99. In each setting, kappas were also excellent (greater than or equal to 0.96). Overall, percent agreement was high for all the PROMIS Depression data elements, ranging from 98 to 99 percent. Percent agreement was also high in each of the four settings, ranging from 97 to 100 percent. There was very little difference in reliability according to setting.

Assessor feedback: Assessors considered PROMIS Depression data elements to be somewhat clinically useful but to have a 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 afterwards. Assessors were also concerned that patients/residents were not being honest about their answers, thus potentially limiting clinical utility. In summary, similar to the PHQ-2 to 9, standardizing depression assessments was important to the assessors in both the survey and focus group discussions, but the relatively high burden of these data elements contributed to a mixed assessment from assessors.

PROMIS Anxiety

The PROMIS Anxiety data elements include eight of a total of 29 data elements from the PROMIS Anxiety Item Bank, which assesses self-reported anxiety-related symptoms. These data elements are not currently used in PAC patient/resident assessment. In the National Beta Test, the PROMIS Anxiety data elements were collected in two versions, which used a reference to “the past 7 days” or “the past 3 days” for symptoms. Images of both versions as tested in the National Beta Test can be found in Volume 5 (www.rand.org/t/RR3004z4).

Results

Assessing anxiety: The “past 7 days” version of PROMIS Anxiety was administered to 1,492 patients/residents. The “past 3 days” version of PROMIS Anxiety was administered to 1,479

patients/residents. Overall, rates were similar between versions and tended to be highest for data elements asking about feeling worried (66 and 63 percent for the “past 7 days” and “past 3 days” versions, respectively) and nervous (56 and 54 percent, respectively) and tended to be lowest for the data element asking about sudden feelings of panic (27 and 25 percent, respectively). Anxiety total scores were also similar on both versions (14.6 and 14.4 percent, respectively). For all data elements and anxiety total score, LTCH patients consistently had higher rates and total scores (on both versions) compared with patients/residents in all other settings.

Missing data: Overall missing data at the data element level for the PROMIS Anxiety ranged from 0.6 percent to 1.81 percent in the “past 7 days” version and 0.41 percent to 1.56 percent in the “past 3 days” version, with minimal setting differences.

Time to complete: The time to complete for the “past 7 days” version and the “past 3 days” versions of PROMIS Anxiety was the same at 2.2 minutes. Time to complete was generally similar across settings; however, for the “past 3 days” version, it took significantly longer to complete the PROMIS Anxiety data elements in HHAs (2.4 minutes) than in IRFs (1.9 minutes) and LTCHs (2.1 minutes) and significantly longer to complete in SNFs (2.2 minutes) than in IRFs (1.9 minutes).

Known groups validity: We hypothesized that anxiety would be associated with gender, toileting, and ability to transfer from lying to sitting. Analyses of these associations using the PROMIS Anxiety t-score supported all three of these hypotheses.

Interrater reliability: Kappas for the PROMIS Anxiety data elements were excellent, ranging from 0.97 to 0.99. In each setting, kappas were also excellent (greater than or equal to 0.95). Overall, percent agreement was high for all the PROMIS Anxiety data elements, ranging from 98 to 99 percent. Percent agreement was also high in each of the four settings, ranging from 97 to 100 percent.

Assessor feedback: Assessors considered PROMIS Anxiety data elements to be somewhat to moderately clinically useful but rated them as having a 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, perhaps because they found them more relevant and were less self-conscious when talking about them.

Results for Data Elements Related to Special Services, Treatments, and Interventions

Four data elements pertaining to Nutritional Approaches and 12 data elements related to Special Treatments were selected for testing in the National Beta Test based on information-gathering activities, expert and stakeholder input, and Alpha feasibility testing. Results of their performance are reviewed below.

Nutritional Approaches

The Nutritional Approaches data elements were Parenteral/IV Feeding, Feeding Tube, Mechanically Altered Diet, and Therapeutic Diet. Parenteral/IV Feeding is a form of feeding in which patients/residents receive nutrition intravenously, bypassing the usual process of eating and digestion. Feeding Tube refers to enteral nutrition, which is the delivery of a nutritionally complete diet directly into the stomach or digestive tract. The Mechanically Altered Diet data element refers to a diet that is specifically prepared to alter the texture or consistency of food to facilitate oral intake. The Therapeutic Diet data element refers to a special diet ordered by a health care practitioner as part of the treatment for a disease or clinical condition. Data elements that indicate therapeutic nutritional approaches are currently collected in each of the instruments for all four PAC settings but vary by assessment instrument. Images of the data elements as tested in the National Beta Test can be found in Volume 6 (www.rand.org/t/RR3004z5).²³

Results

Assessing nutritional approaches: In the National Beta Test, the Nutritional Approaches data elements were administered to 2,926 patients/residents across settings. Overall, three of the four Nutritional Approaches were seldom performed for individuals in the admission sample (ranging from 1 to 10 percent overall). However, approximately half of patients/residents were receiving a therapeutic diet (52 percent), and this was true across all settings, although rates were somewhat higher in LTCHs (59 percent) compared with other settings (49 to 54 percent).

Missing data: Overall missing data at the data element level for the Nutritional Approaches ranged from 0.7 to 1.3 percent, with minimal setting differences.

Time to complete: The average time to complete the Nutritional Approaches data elements was 0.9 minutes overall, with minimal setting differences.

Known groups validity: We hypothesized that having a mechanically altered diet would be associated with age, stroke, toileting, and ability to transfer from lying to sitting. Analyses of these associations supported three of the four hypotheses.

Interrater reliability: Kappa interrater reliability coefficients, overall, were not calculated for two of the four Nutritional Approaches data elements (Parenteral/IV Feeding and Feeding Tube), given that prevalence rates were out of range for stable and interpretable kappa estimates as determined by study power calculations. However, overall kappas were good for mechanically altered diet (0.65) and therapeutic diet (0.60) data elements. At the setting level, kappa was also good for mechanically altered diet in LTCHs (0.69) and SNFs (0.70), moderate in IRFs (0.53), and not calculated in HHAs because of low prevalence rates; for therapeutic diet, kappa was good in IRFs (0.70), LTCHs (0.62), and SNFs (0.61) and moderate in HHAs (0.43). Overall percent agreement was high for all Nutritional Approaches data elements, ranging from 80 to

²³ Edelen et al., 2019d.

100 percent. With one exception, percent agreement was high in each of the four settings, ranging from 80 to 100 percent. In HHAs, percent agreement for therapeutic diet was 71 percent.

Assessor feedback: Assessors considered the Nutritional Approaches data elements to be important for conveying patient/resident healthcare needs, complexity, and progress, but rated data collection of these data elements to be moderately burdensome, as information on nutritional therapies was sometimes difficult to locate in the medical record.

Special Treatments

Data elements on Special Treatments (chemotherapy [IV, oral, other], radiation, oxygen therapy [intermittent, continuous, high-concentration oxygen delivery system], suctioning [scheduled, as needed], tracheostomy care, invasive mechanical ventilator, non-invasive mechanical ventilator [BiPAP, CPAP], IV medications [antibiotics, anticoagulation, other], transfusions, dialysis [hemodialysis, peritoneal dialysis], and IV access [peripheral IV, midline, central line, other]) were assessed by reviewing the patient's/resident's medical record. With the exception of IV access and the sub-data elements, similar data elements are currently collected in the MDS. In addition, similar versions of invasive mechanical ventilator, non-invasive mechanical ventilator, IV medications, and dialysis are currently collected in the LCDS. Images of the data elements as tested in the National Beta Test can be found in Volume 6 (www.rand.org/t/RR3004z5).

Results

Assessing special treatments: In the National Beta Test, the Special Treatments data elements were administered to 2,926 patients/residents across settings. For the Special Treatments data elements, the majority of treatments were not performed for individuals in the admission sample, and those that were performed tended to be more common in the LTCH setting. Specifically, oxygen therapy was administered to 20 percent of the overall sample and 44 percent of LTCH patients. Similarly, IV medication and IV access were noted among 25 percent and 24 percent of the overall sample, respectively, but in 77 percent and 91 percent of LTCH patients.

Missing data: Overall missing data at the data element level for Special Treatments ranged from 0.5 to 1.4 percent overall, with minimal setting differences.

Time to complete: The average time to complete the Special Treatments data elements was 2.3 minutes overall, with minimal setting differences. Across settings, the average time per data element was approximately 13 seconds.

Known groups validity: We hypothesized that IV access would be associated with having sepsis. Analysis of this association supported this hypothesis.

Interrater reliability: Kappas for Special Treatments ranged considerably. For oxygen therapy data elements, kappas ranged from moderate to excellent (0.55 to 0.82) overall and ranged from good to excellent (0.68 to 0.86) across settings, with one exception in LTCHs (0.35: continuous oxygen therapy). IV medication kappas, overall, ranged from moderate to excellent

(0.46 to 0.88), with the exception of the kappa for anticoagulation, which was poor (0.13). IV medication kappas ranged from moderate to excellent (0.46 to 0.84) across settings, with the exception of kappa for IV medications in HHAs (0.15) and anticoagulants in LTCHs (0.13), which were poor. IV access kappas were excellent, ranging from 0.81 to 0.90 overall and from good to excellent (0.74 to 0.81) across settings. Kappa interrater reliability coefficients, overall, are not reported for remaining data elements (e.g., cancer treatments) given that prevalence rates were out of range for stable and interpretable kappa estimates as determined by study power calculations. Overall percent agreement was high for all Special Treatments data elements, ranging from 88 to 100 percent. Percent agreement was also high in each of the four settings, ranging from 83 to 100 percent.

Assessor feedback: Assessors considered the Special Treatments data elements to be important for clinical care planning but also noted moderate to high burden for these data element because of difficulty at times locating the information in the chart.

Results for Data Elements Related to Medical Conditions: Pain

Several patient interview data elements were selected to assess pain (collectively referred to as the Pain Interview) for testing in the National Beta Test based on information-gathering activities, expert and stakeholder input, and Alpha feasibility testing. Results of their performance are reviewed below.

Pain Interview

The Pain Interview data elements assess the presence of pain and pain frequency, along with three data elements evaluating for (1) pain interference with sleep, therapy activities, and other activities; (2) pain severity; and (3) relief from pain due to pain treatments or medications. These data elements were completed through patient/resident interview. Similar versions of some of the Pain data elements are included in the OASIS and the MDS. In the National Beta Test, the Pain Interview data elements were collected in two versions. One version asked about pain experiences in “the past 3 days,” and the other version asked about pain in “the past 5 days.” Images of both versions as tested in the National Beta Test can be found in Volume 5 (www.rand.org/t/RR3004z4).

Results

Assessing pain: The version of the pain interview for the “past 3 days” was administered to 1,520 patients/residents. The version for the “past 5 days” was administered to 1,511 patients/residents. Overall, at least three-quarters of patients/residents indicated pain presence (“past 3 days” version: 75 percent, “past 5 days” version: 80 percent). Generally, for each setting, indication of pain presence was slightly higher for the “past 5 days” version, except in IRFs, where rates were slightly higher for the “past 3 days” version. Further, nearly two-thirds of

patients/residents indicated that pain interfered with sleep (“past 3 days” version: 67 percent, “past 5 days” version: 64 percent). In each setting, indication of pain interfering with sleep was slightly higher for the “past 3 days” version, except in SNFs, where rates were slightly higher for the “past 5 days” version. Moreover, patients/residents in LTCHs tended to report greater pain interference with sleep (on both versions) than patients/residents in other settings.

Missing data: Missing data did not exceed 2.8 percent for either version of the Pain Interview, with minimal setting differences.

Time to complete: For all patients/residents who completed the Pain Interview data elements, time to complete was an average of 2.6 minutes. For patients/residents reporting no pain ($n = 440$), and who therefore ended the Pain Interview after stating they had no pain, the average time to complete was 1.2 minutes. For patients/residents reporting pain ($n = 1331$), the average time to complete was 3.0 minutes. There were minimal setting differences for time to complete the Pain Interview data elements.

Known groups validity: We hypothesized that presence of pain would be associated with gender, age, toileting, and ability to transfer from lying to sitting. Analyses of these associations supported all four of these hypotheses.

Interrater reliability: 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 data elements were excellent, ranging from 0.96 to 0.98. At the setting level, kappas were also excellent (greater than or equal to 0.83). Percent agreement, overall, was excellent, ranging from 96 to 99 percent. Percent agreement was also high across the four settings, ranging from 95 to 100 percent.

Assessor feedback: Feedback was strongly positive for the Pain Interview data elements, especially the questions about pain’s impact on function. Facility/agency 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. Assessors agreed that the Pain Interview data elements had high clinical utility, particularly for patient/resident transfer and care planning, but they also noted that patients/residents experienced difficulty with recalling and characterizing pain experienced in the past.

Results for Data Elements Related to Impairments

Several data elements were selected to assess sensory and continence impairments for testing in the National Beta Test based on information-gathering activities, expert and stakeholder input, and Alpha feasibility testing. Results of their performance are reviewed below.

Hearing and Vision

The data elements Ability to Hear and Ability to See assess the level of hearing and vision impairment and consist of one question each. Ability to Hear is currently assessed in the MDS,

and Ability to See is currently assessed in the MDS and OASIS. An image of the data elements as tested in the National Beta Test can be found in Volume 6 (www.rand.org/t/RR3004z5).

Results

Assessing sensory impairment: The data elements to assess sensory impairment were administered to 3,065 patients/residents. Overall, 74 percent of patients/residents had adequate hearing, and 17 percent had minimal difficulty hearing. Rates for patients/residents with adequate hearing were slightly higher in LTCHs (81 percent) and lower in HHAs (65 percent). Overall, 78 percent of patients/residents had adequate vision, and 16 percent had impaired vision. Rates for patients/residents with adequate hearing were slightly higher in IRFs (85 percent) and lower in HHAs (73 percent).

Missing data: Among those who were administered the sensory impairment data elements, missing data was 0.3 percent for hearing and 0.6 percent for vision overall, with minimal setting differences.

Time to complete: On average, the two sensory impairment data elements took 0.6 minutes to complete combined, with minimal differences across settings.

Known groups validity: We hypothesized that hearing and vision impairment would be associated with age, stroke, and needing assistance with toileting. In addition, we hypothesized that hearing impairment would be associated with gender. Analyses of these associations supported two of the four hypotheses for hearing and two of the three hypotheses for vision.

Interrater reliability: Kappas were computed on 960 patients/residents who were assessed by research nurses and facility/agency staff assessor pairs: 197 in HHAs, 258 in IRFs, 237 in LTCHs, and 268 in SNFs. The overall kappa for the Ability to Hear was good at 0.65 and moderate to good across settings (0.58 to 0.71). The overall kappa for the Ability to See was moderate at 0.56 and moderate to good across settings (0.47 to 0.67). Percent agreement was moderate for these data elements (84 percent for hearing, 83 percent for vision). Across settings, percent agreement for Ability to Hear ranged from 83 to 87 percent, and percent agreement for Ability to See ranged from 75 to 90 percent (highest in IRFs).

Assessor feedback: Findings from both survey and focus groups confirmed that vision and hearing are important, clinically relevant concepts to assess according to facility/agency staff and research nurses, especially as the patient/resident is first entering their facility's/agency's care. These data elements posed a low burden because of their observational nature. However, the data elements tended to be collected in slightly different ways (e.g., for hearing, by noting how high the patient/resident keeps the volume on the TV versus by determining whether the patient/resident can hear the assessor if she or he speaks in a soft tone) based on assessor preferences and experience, which might compromise standardized assessment.

Continence Interview

The Continence Interview data elements assess the occurrence of incontinent events as well as the patient's/resident's perception of the burden of bladder and bowel incontinent events. These data elements are completed through patient/resident interview and are not assessed in any of the current PAC assessment instruments. An image of the data elements as tested in the National Beta Test can be found in Volume 6 (www.rand.org/t/RR3004z5).

Results

Assessing bladder and bowel incontinent events using patient interview: The Continence Interview data elements were administered to 2,977 patients/residents. Overall, more than one-third of patients/residents reported any bladder incontinent events, and 20 percent reported bowel incontinent events. Rates were generally similar across settings. However, bladder incontinent events were slightly higher in SNFs (42 percent), and bowel incontinent events were slightly higher in LTCHs (29 percent). Most events were reported to have posed at least a small problem (only 13 percent of bladder and 10 percent of bowel incontinent events were rated as being “no problem”).

Missing data: Missing data for the Continence Interview data elements ranged from 0.2 to 0.7 percent overall, with minimal setting differences. Further, overall rates for “declined to respond” ranged from 0.03 to 0.17 percent, with minimal setting differences.

Time to complete: On average, the Continence Interview data elements took 1.4 minutes to complete, with minimal setting differences.

Known groups validity: We hypothesized that bladder incontinent events would be associated with gender, age, length of stay, stroke, toileting, and ability to transfer from lying to sitting. Analyses of these associations using whether bladder incontinent events were reported in the Continence Interview supported five of the six hypotheses.

Interrater reliability: For the Continence Interview data elements, kappas were computed on the 927 patients/residents who were assessed by research nurse and facility/agency staff assessor pairs. Overall kappas for the Continence Interview data elements were excellent, ranging from 0.96 to 0.98. In each of the four settings, kappas were also excellent (greater than or equal to 0.94). Percent agreement, overall, was also excellent, ranging from 98 to 99 percent, and was similarly high across the four settings, ranging from 96 to 100 percent.

Assessor feedback: Facility/agency staff considered the Continence Interview to be clinically relevant; they explained that it was useful to know how incontinence could be affecting the patient's/resident's functionality and quality of life. However, a weakness of this data element noted by both types of assessors was the accuracy of self-reporting, as patients'/residents' interview responses sometimes contradicted the medical record. Patients/residents who reported no incontinence would sometimes contradict their chart, a spouse or family member's report (if one was present), or physical evidence observed by the assessor.

Continence Chart Review

The Continence Chart Review data elements documented the use of equipment and appliances to manage incontinence, whether they were placed in the current care setting, the primary reason for catheter placement, the need for assistance to manage the use of appliances, and the frequency of incontinent events. Similar data elements for Appliance Use are currently used in the OASIS and MDS, and for frequency of events in the OASIS, IRF-PAI, LCDS, and MDS. An image of the data elements as tested in the National Beta Test can be found in Volume 6 (www.rand.org/t/RR3004z5).

Results

Assessing bladder and bowel incontinent events using chart review: The Continence Chart Review data elements were administered to 2,926 patients/residents. The use of a bladder appliance was uncommon, with an indwelling urethral catheter used most frequently (10 percent overall) and most commonly used among LTCH patients (34 percent). Of the small percentage of appliances placed in the current setting (20 percent), retention was the most common reason for placement (8 percent), and most patients/residents with a bladder appliance needed some assistance with management (89 percent overall). Slightly less than 40 percent of the overall sample was noted as having bladder incontinent events, with the fewest in IRFs (26 percent).

Bowel incontinent events were noted less frequently, at 14 percent; however, they were more common in LTCHs (26 percent). Bowel appliance use was infrequent overall (5 percent) but higher in LTCHs (12 percent) and rarely placed in the current setting. Most patients/residents with a bowel appliance needed assistance with its management (87 percent).

Missing data: Missing data for the Continence Chart Review data elements ranged from 0.0 to 6.9 percent overall, with minimal setting differences but with one exception. Missing data were higher for the “bladder incontinent event” data element in HHAs (20 percent) than in any other setting (0.7 to 6.1 percent).

Time to complete: On average, the Continence Chart Review data elements took 3.5 minutes to complete (approximately 21 seconds per data element), with minimal setting differences.

Known groups validity: We hypothesized that bladder appliance use would be associated with gender, toileting, and ability to transfer from lying to sitting. Analyses of these associations supported all three of the hypotheses.

Interrater reliability: For the Continence Chart Review data elements, paired assessments were completed on 884 patients/residents who were assessed by research nurse and facility/agency staff assessor pairs. Kappas for many of the Continence Chart Review data elements were not calculated because of proportions of responses being out of range to support stable kappa estimates. The kappas that were calculable (i.e., for data elements observed at higher frequencies) tended to be substantial/good overall (0.66–0.79). Overall percent agreement was moderate to excellent, ranging from 74 to 100 percent. This general range was also observed in LTCHs (77 to 99 percent) and SNFs (71 to 100 percent); however, the range was wider in

HHAs (50 to 100 percent) and IRFs (67 to 100 percent). Data elements with lower percent agreement in HHAs and IRFs were those asking about whether the bowel appliance was placed in the current setting and whether the patient needs help with management.

Assessor feedback: Facility/agency staff and research nurses rated continence among the top five data elements in terms of clinical utility in the assessor survey. In the focus groups, facility/agency staff explained that continence was very clinically relevant for care decisionmaking and planning, specifically for determining interventions, planning discharge, and protecting skin integrity. However, assessors also reported that continence data elements that required chart review, even with an electronic medical record, were problematic because of high data collection burden. This was especially true for the home health setting.

Results for Data Elements Related to Other Clinical Categories

Several data elements were selected to assess other clinical categories for testing in the National Beta Test, including data elements related to global health, care preferences, and medication reconciliation, based on information-gathering activities, expert and stakeholder input, and Alpha feasibility testing. Results of their performance are reviewed below.

PROMIS Global Health

The PROMIS Global Health profile assessment consists of ten data elements that address health-related quality of life, including functioning and well-being in physical, mental, and social domains of life. PROMIS Global Health is intended to globally reflect the patients'/residents' assessment of their health. These data elements are completed through patient/resident interview and are not currently in use in PAC assessment. An image of the data elements as tested in the National Beta Test can be found in Volume 7 (www.rand.org/t/RR3004z6).²⁴

Results

Assessing global health: The standard version of the PROMIS Global Health data elements, which ask about experience either “in general” or “in the past 7 days,” was administered to 1,528 patients/residents. A modified version asking about experience “in the past 3 days” was administered to 1,521 patients/residents. Global Physical Health total scores were similar on both versions (11.5 and 11.2 for the “in general/past 7 days” and “past 3 days” versions, respectively) but were lowest in LTCHs (10.8 and 10.4, respectively). Global Mental Health total scores were also similar between versions but were slightly higher for the “in general/past 7 days” version (13.4) compared with the “past 3 days” version (12.5). This general trend was also observed across the four settings, with scores on both versions being slightly lower in LTCHs (12.3 and 11.8, respectively) than other settings.

²⁴ Edelen et al., 2019e.

Missing data: Overall missing data at the data element level for the PROMIS Global Health data elements ranged from 0.72 percent to 4.21 percent in the two versions, with minimal setting differences.

Time to complete: On average, the time to complete for the two versions of the PROMIS Global Health data elements (“in general/past 7 days” and “past 3 days”) was 3.5 and 3.7 minutes, respectively. Time to complete, on average, was generally similar across settings. However, the “past 3 days” version took significantly less time to complete in IRFs (2.8 minutes) than in any other setting (3.6 to 3.9 minutes).

Known groups validity: We hypothesized that PROMIS Global Physical Health t-score and PROMIS Global Mental Health t-score would be associated with toileting and ability to transfer from lying to sitting. Analyses of these associations supported the two hypotheses for both t-scores.

Interrater reliability: Kappas for the PROMIS Global Health data elements were excellent for both versions, ranging from 0.89 to 1.00. In each of the four settings, kappas were also excellent (greater than or equal to 0.89) for both versions. Overall percent agreement was high, ranging from 95 to 99 percent, and percent agreement was similarly high in each of the four settings, ranging from 91 to 100 percent.

Assessor feedback: Assessment of a patient’s/resident’s perception of their health was important to assessors. Assessors considered some of the PROMIS Global Health data elements to be clinically useful but also considered many of them to be not relevant for the majority of PAC patients/residents (e.g., data elements asking about ability to carry out daily activities, such as walking and shopping).

Care Preferences

The Care Preferences data elements assessed patient/resident preferences for involvement of family/friends in care decisions and patient/resident involvement in their own care, as well as whether the patient/resident has a designated health care agent. These data elements are completed through patient/resident interview or medical record review. A similar data element is included in the MDS that assesses how important it is to the patient/resident to have his or her family involved in discussions about care. An image of the data elements as tested in the National Beta Test can be found in Volume 7 (www.rand.org/t/RR3004z6).

Results

Assessing care preferences: The Care Preference interview data elements were administered to 2,980 patients/residents, and the designated Health Care Agent data element was completed for 2,923 patients/residents. Overall, 73 percent of patients/residents indicated that it was *very important* to have a close person involved in their care. This was generally consistent across settings but was slightly higher in IRFs (76 percent) compared with other settings (72 to 73 percent). Further, 88 percent of patients/residents overall preferred *to know as much as they*

could about their condition, which was generally similar across settings. Overall, a health care agent was noted for 30 percent of patients but was less frequently noted in LTCHs (22 percent) compared with all other settings (28 to 34 percent).

Missing data: For the Care Preference interview data elements, overall missing data ranged from 0.70 to 0.94 percent, with minimal setting differences. For the data element for a designated health care agent, although there was technically no missing data, the large proportion of patients/residents for whom there was no designated health care agent includes patients/residents for whom this information was unknown or unable to be assessed.

Time to complete: On average, the two Care Preference interview data elements took 1.5 minutes to complete but took slightly less time in IRFs (1.3 minutes) than in all other settings (1.5 minutes). The chart review data element, on average, took 1.2 minutes to complete overall, with minimal setting differences.

Known groups validity: We hypothesized that preferences about knowing details of condition and treatment would be associated with age and gender. Analyses of these associations supported both of these hypotheses.

Interrater reliability: For Care Preferences interview data elements, kappas were computed on 930 patients/residents who were assessed by research nurses and facility/agency staff assessor pairs. Kappa and percent agreement for *importance of a close person involved in care* were excellent overall (0.96, 98 percent, respectively) and in each of the four settings (0.93 to 0.98, 98 percent, respectively). Similarly, kappa and percent agreement for *how much a patient/resident prefers to know about their condition* were excellent overall (0.96, 99 percent, respectively) and in each of the four settings (0.94 to 0.98, 99 to 100 percent, respectively).

For the Care Preference chart review data element, kappas were computed on 881 patients/residents. Overall kappa for whether a patient/resident had a designated health care agent was moderate (0.56) overall and moderate to good (0.51 to 0.63) in each of the four settings. Percent agreement overall was 83 percent and ranged from 79 to 87 percent across the four settings.

Assessor feedback: The Care Preferences data elements were deemed important by both facility/agency staff and research nurses. Assessors also considered the Care Preferences set overall to be among the least burdensome data elements but indicated higher burden related to the data elements about a health care agent.

Medication Reconciliation

The Medication Reconciliation data elements assess whether a patient/resident was taking any medications in several classes, whether the original prescriber for each medication noted the indication, whether there were discrepancies among a patient's/resident's multiple medication lists and whether the discrepancies were addressed by involving the patient/resident or caregiver, whether discrepancies were communicated to a physician within 24 hours, whether recommended actions were carried out within 24 hours, and whether the reconciled medication

list was communicated to the patient/resident, care team, and pharmacy. Similar data elements are currently collected in all four PAC instruments and assess whether significant medication issues were identified and whether a physician was contacted and recommended actions were completed. In addition, the MDS currently collects resident medications received by classification. An image of the data elements as tested in the National Beta Test can be found in Volume 7 (www.rand.org/t/RR3004z6).

Results

Assessing medication reconciliation: The Medication Reconciliation data elements were administered to 2,951 patients/residents in the admission sample: 627 in HHAs, 769 in IRFs, 459 in LTCHs, and 1,096 in SNFs. Overall, medication classes taken ranged from 12 percent (antipsychotics) to 51 percent (opioids). Opioids were the most common medication class in HHAs and SNFs and second most common in IRFs and LTCHs, where anticoagulants were most common. On average, the total number of medication classes taken was 1.83; however, this varied by setting. The average number of medication classes taken was higher in LTCHs (2.78) compared with all other settings (1.29 to 1.88). Further, the average number of medication classes taken in HHAs (1.29) was lower than in IRFs (1.88) and SNFs (1.71). Indications for medication classes taken, overall, ranged from 45 percent (anticoagulants and antiplatelets) to 92 percent (opioids) and were consistently highest for opioids across all settings. Discrepancies involving medication classes taken, overall, ranged from 3 percent (antiplatelets) to 6 percent (antimicrobials) and were consistently highest for antimicrobials across settings, except in SNFs, where discrepancies were highest for antipsychotics (6 percent). Whether the *reconciled medication list was communicated*, overall, ranged from 7 percent (communicated to none) to 83 percent (communicated to prescribers) and was generally highest for “communicated to prescribers” across settings, except in HHAs, where “communicated to patient/caregiver” was highest.

Missing data: Missing data for the Medication Reconciliation data elements ranged from 0.0 to 4.2 percent, with minimal setting differences.

Time to complete: On average, the Medication Reconciliation data elements took 3.2 minutes to complete overall and took less time to complete in HHAs (2.9 minutes) than in IRFs (3.4 minutes) and LTCHs (3.5 minutes).

Known groups validity: Although we did not have hypotheses related to the number of medication classes taken by the patient/resident, we found significant associations with gender, age, disposition at discharge, sepsis, level of assistance needed with toileting, and ability to transfer from lying to sitting.

Interrater reliability: For the Medication Reconciliation data elements, kappas and percent agreement were computed on 900 patients/residents who were assessed by research nurses and facility/agency staff assessor pairs. Kappas for medication classes taken were good to excellent (0.72 to 0.89) overall and in each of the four settings (0.71 to 0.97). Percent agreement was also

high, ranging from 92 to 95 percent overall and from 91 to 99 percent across the four settings. Overall kappas for indication data elements were good to excellent (0.65 to 0.87) overall and moderate to excellent (0.54 to 1.00) across settings, except for hypoglycemics (0.39) and antipsychotics (0.33) in HHAs, which were fair. Percent agreement ranged from 82 to 94 percent overall and 77 to 100 percent across the four settings but was lower for hypoglycemics (69 percent) and antipsychotics (63 percent) in HHAs. Kappa for recommended discrepancy action carried out within 24 hours was excellent overall (0.85) and good to excellent (0.69 to 1.00) in each of the four settings. Percent agreement was 82 percent overall and ranged from 73 to 100 percent across the four settings. Overall kappas for reconciled medication lists being communicated data elements were moderate (0.42 to 0.57) overall and fair to good (0.26 to 0.66) in each of the four settings. Percent agreement ranged from 79 to 92 percent overall and 78 to 99 percent across the four settings, except for list communicated to prescribers in HHAs (69 percent) and list communicated to pharmacy in LTCHs (71 percent).

Assessor feedback: Assessors considered the Medication Reconciliation data elements to be moderately to extremely clinically useful. There was mixed feedback regarding assessment burden. Overall assessment burden was reported to be high because of confusion about what constituted a discrepancy, lack of documentation (e.g., communicating discrepancies and follow-up), and differences among settings. However, it was noted that medication reconciliation was much easier in facilities that had electronic medical record software that already incorporated this process and was also easier for those familiar with electronic medical records and other documentation systems in use.

Results for Data Elements Related to Observational Assessments of Cognitive Function, Mental Status, and Pain

Several data elements were selected specifically for assessment among patients/residents who are unable to communicate for testing in the National Beta Test based on information-gathering activities, expert and stakeholder input, and Alpha feasibility testing. These data elements assessed the clinical categories of cognitive function, mental status, and pain. Results of their performance are reviewed below.

Staff Assessment of Mental Status

The Staff Assessment of Mental Status assesses aspects of cognitive status, including long-term memory, short-term memory, memory/recall ability, and decisionmaking based on staff observation, information provided by staff and family/friends, and medical records. These data elements are intended for use among patients/residents in all PAC settings who are unable to complete the interview-administered BIMS because of refusal, nonsensical answers, or inability to make themselves understood at least some of the time. The Staff Assessment of Mental Status

is currently collected in the MDS and IRF-PAI. An image of the data elements as tested in the National Beta Test can be found in Volume 8 (www.rand.org/t/RR3004z7).²⁵

Results

Assessing cognitive function by observation: The Staff Assessment of Mental Status was completed for 513 patients/residents. Overall, 84 percent of patients/residents who were assessed had a short-term memory problem and 76 percent had a long-term memory problem. Further, 76 percent of patients/residents were severely impaired in their ability to make everyday decisions, with higher rates in LTCHs (82 percent) and SNFs (81 percent) compared with HHAs (67 percent) and IRFs (57 percent).

Missing data: Missing data for the Staff Assessment of Mental Status data elements ranged from 2.9 to 33.5 percent, the majority of which reflected cases where responses were unknown or unable to be assessed. However, excluding the one data element with 2.9 percent missing (ability to make decisions regarding everyday tasks), the missing data ranged from 23.4 to 33.5 percent. Missing data rates were similar for HHAs, IRFs, and SNFs but were slightly higher in LTCHs.

Time to complete: Overall, the Staff Assessment of Mental Status data elements required, on average, 2.6 minutes to complete, with minimal setting differences.

Known groups validity: Because of the heterogeneous nature of the non-communicative population, we did not form hypotheses about the associations between cognitive impairment and patient/resident characteristics and clinical conditions. In the analysis, we found that impaired ability to make decisions assessed in the Staff Assessment of Mental Status was associated with length of stay, disposition at discharge, and activities of daily living.

Interrater reliability: Kappa coefficients for the Staff Assessment of Mental Status data elements were good to excellent (0.74 to 0.94) overall and good to excellent (0.64 to 1.00) in each of the four settings. Percent agreement was also high, ranging from 93 to 98 percent overall and 89 to 100 percent across the four settings.

Assessor feedback: According to the assessor survey, facility/agency staff and research nurses found the Staff Assessment of Mental Status to be somewhat to moderately clinically useful. When asked about burden in the assessor survey, on average, facility/agency staff thought that it was “slightly difficult” to collect information, and the data element scored in the middle range on burden, relative to other data elements. In contrast, research nurses rated this data element as one of the most burdensome to collect because it was difficult to complete for assessors with only limited familiarity with the patient/resident.

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

The Patient Health Questionnaire-9 Observational Version (PHQ-9-OV) assesses depressed mood in patients/residents who cannot complete a patient/resident mood interview because of an

²⁵ Edelen et al., 2019f.

inability to communicate. These data elements are completed through interviews with staff, family members, and/or other caregivers who know the patient/resident best and by reviewing medical records. The PHQ-9-OV is currently used in the MDS. An image of the data elements as tested in the National Beta Test can be found in Volume 8 (www.rand.org/t/RR3004z7).

Results

Assessing mental status by observation: The PHQ-9-OV was completed for 501 patients/residents. Overall scores on the PHQ-9-OV indicate that 63 percent of patients/residents had no depression, 22 percent had minor depression, and 15 percent had major depression. At the setting level, a higher proportion of patients/residents in IRFs (25 percent) had major depression compared with all other settings (10 to 13 percent). Overall, the average PHQ-9-OV total score was 6.6 and was highest in LTCHs (7.8) and IRFs (7.1) compared with HHAs (5.5) and SNFs (5.8).

Missing data: Missing data for the PHQ-9-OV ranged from 12.4 to 44.3 percent, the majority of which reflected cases in which responses were unknown or unable to be assessed. Missing data rates were similar among HHAs, IRFs, and SNFs and were slightly higher in LTCHs.

Time to complete: On average, the time to complete the PHQ-9-OV was 3.5 minutes overall and took significantly longer to complete in HHAs (4.5 minutes) compared with all other settings (3.3 to 3.7 minutes).

Known groups validity: Because of the heterogeneous nature of the non-communicative population, we did not form hypotheses about the associations between the PHQ-9-OV and patient/resident characteristics and clinical conditions. In the analysis, across all patient/resident characteristics and clinical conditions, there were not significant associations with the patient's/resident's PHQ-9-OV total scores.

Interrater reliability: Kappa coefficients for the PHQ-9-OV were computed based on 487 total patients/residents. Kappa coefficients were excellent overall, ranging from 0.85 to 0.98. In each of the four settings, kappas were also excellent, ranging from 0.84 to 1.00. Percent agreement was high, ranging from 96 to 99 percent overall and 93 to 100 percent across the four settings.

Assessor feedback: Facility/agency staff and research nurses considered standardized assessment of signs and symptoms of depression to be important. According to the assessor survey, facility/agency staff and research nurses rated the Staff Assessment of Patient/Resident Mood as somewhat to moderately clinically useful and as having a higher data collection burden than many of the other data elements.

Observational Assessment of Pain or Distress

The data elements that make up Observational Assessment of Pain or Distress collect staff observations of patients'/residents' expressed behavioral indicators of potential pain or distress and were administered to all patients/residents who are unable to communicate (i.e., who cannot

reliably make themselves understood via verbal communication, written communication, communication board, or eye blinks). These data elements were completed through interviews with staff, family members, and/or other caregivers; observation of the patient/resident during care activities; and review of the medical record. A similar data element is currently included in the MDS. An image of the data elements as tested in the National Beta Test can be found in Volume 8 (www.rand.org/t/RR3004z7).

Results

Assessing pain by observation: The Observational Assessment of Pain or Distress was completed for 545 patients/residents. The Observational Assessment of Pain or Distress results overall show that for 56 percent of patients/residents, at least one pain or distress sign was observed and documented. However, this rate was higher in HHAs (69 percent) and LTCHs (64 percent) than in LTCHs and SNFs, where approximately half of all patients/residents had pain or distress signs observed or documented.

Missing data: Missing data for the Observational Assessment of Pain or Distress data elements ranged from 0.0 to 4.4 percent, all of which reflected cases where responses were unknown or unable to be assessed. Missing data were similar across settings.

Time to complete: On average, the time to complete the Observational Assessment of Pain or Distress was 2.4 minutes, with minimal setting differences.

Known groups validity: Because of the heterogeneous nature of the non-communicative population, we did not form hypotheses about the associations between observations of any signs of pain and patient/resident characteristics and clinical conditions. In the analysis, across all patient/resident characteristics and clinical conditions, there were not significant associations with observed signs of pain.

Interrater reliability: Kappa coefficients for the Observational Assessment of Pain or Distress were excellent overall, ranging from 0.81 to 0.90, and were good to excellent (0.65 to 1.00) in each of the four settings. Percent agreement was high, ranging from 89 to 98 percent overall and 83 to 100 percent across the four settings.

Assessor feedback: Facility/agency staff and research nurses considered pain to be a clinically useful measurement across PAC settings. According to the assessor survey, facility/agency staff and research nurses found the observational assessment of pain or distress to be somewhat to moderately clinically useful and thought that it was only slightly difficult to collect information for this data element.

Additional Results Related to Evaluation of SPADE Validity

Day 3, 5, 7 Repeat Assessment

A subset of candidate SPADEs from the communicative admission assessment who are assessed via patient/resident interview or other interaction were assessed repeatedly by the same

assessor on the same patient/resident on Days 3, 5, and 7. The SPADEs included in the Day 3, 5, 7 repeat assessment were BIMS, CAM, Behavioral Signs and Symptoms, Expression and Understanding, and Pain Interview. The goal of the Day 3, 5, 7 repeat assessment was to understand whether it matters which day patients/residents are assessed (e.g., at Day 3, Day 5, or Day 7). For all but Pain Interview, there was little variability in scores across assessment days, and thus it did not really matter on which day the data elements were administered. For Pain interview, however, we saw a general improvement across the repeated assessments days, suggesting the possible need for frequent, even daily, assessment of pain, which is commonly done in clinical practice.

Chart Review at Admission Versus Day 3, 5, and 7 and at Discharge Versus Two Days Prior to Discharge

Two candidate SPADE sets that are assessed primarily via chart review, Continence Chart and SSTIs, were assessed over different look-back time frames to determine whether rates of occurrence depend on the look-back period and, if so, identify the earliest chart review day that captures the majority of occurrences. There were two look-back designs evaluated: (1) chart review on admission versus Day 3, 5, and 7 and (2) chart review at discharge versus two days prior to discharge. Data elements in the chart review look-back design included SSTIs (i.e., Nutritional Approaches and Special Treatments) and Continence. Results showed that if an appliance, service, or treatment was present for a given patient/resident, it tended to be noted on admission (Day 1) and on the day of discharge. That is, assessors gained very little additional information about these data elements when extending the chart review beyond those days, implying that assessment of these SPADEs will yield similar results regardless of the actual day of the coding.

Stability/Change from Admission to Discharge

All communicative candidate SPADEs were administered at both admission and discharge in compliance with the requirements of the IMPACT Act. Data from these two assessments were used to evaluate feasibility of data collection at the two points and the extent of stability or change over time from admission to discharge. For most data elements, with a few exceptions, responses and scores were generally stable from admission to discharge. However, for Pain Interview, PHQ-2 to 9, PROMIS Depression, PROMIS Anxiety, and PROMIS Global Health, scores tended to improve over the course of the PAC stay. This improvement in symptoms at discharge implies that assessment of symptoms of depression, anxiety, pain, and physical and mental health may be most informative at both admission and discharge to obtain a complete picture of a patient's/resident's mental and physical status and pain during his or her PAC stay.

4. Discussion

The goal of the National Beta Test was to evaluate the performance of a set of candidate SPADEs from several clinical categories delineated in the IMPACT Act. The National Beta Test was developed to help determine candidate data element suitability and to identify the data elements that would be clinically useful and reliable for cross-setting standardization. Several aspects of the work prior to and during the National Beta Test provided a solid basis for interpreting results from this field test and having strong confidence in the conclusions made from these data.

Prior to National Beta Testing, the RAND team and its subcontractors, working closely with CMS, conducted rigorous information-gathering and pilot testing phases and simultaneously maintained significant interaction with stakeholders through numerous stakeholder engagement opportunities. This rigorous process ensured that the candidate SPADEs included in the National Beta Test have the most potential to meet the mandates of the IMPACT Act.

The National Beta Test design was well planned, well vetted, and sound, enabling evaluation of feasibility and psychometric performance of the candidate SPADEs across settings and within each setting type. Further, the design included features to evaluate different look-back periods for subsets of SPADEs, as well as stability and/or change from admission to discharge. Finally, we were able to merge the National Beta Test assessment data with current assessment data (i.e., the OASIS, IRF-PAI, LCDS and MDS), which allowed for preliminary evaluation of the validity of the candidate SPADEs.

The implementation of the field testing was also quite rigorous. The data collectors participated in thorough in-person training at the start of the field period and had ongoing support throughout the field period. Further, the data collection progress was closely monitored, data quality was ensured with multiple checks and balances so that the data were of the utmost quality, and assessment completion rates were optimized with respect to the ability of participating facilities and agencies to meet target assessment goals.

PAC facilities/agencies participating in the National Beta Test were fairly comparable to the national population of PAC facilities/agencies, and all setting types were well represented with sufficient geographic spread. Importantly, comparative analyses indicated that the National Beta Test patient/resident participant sample was closely aligned with the national population of PAC patients and residents in most respects.

Summary Evaluation of Candidate SPADEs

Throughout all phases of this project, we have maintained consideration of four critical properties in evaluation of candidate SPADEs:

- feasibility for use in PAC
- psychometric performance
- potential for improving quality
- utility for describing case mix.

In what follows, the performance of the candidate SPADEs is summarized according to these principles. A summary of the results from the National Beta Test for each candidate SPADE set is presented in Table 4.1.

Feasibility for PAC Cross-Setting Use

Results of the National Beta Test provided strong support for the feasibility of use of the majority of candidate SPADEs. We evaluated feasibility in terms of missing data, time to complete, and perceived assessment burden. In almost all cases, feasibility results according to these specifications were similar across the four setting types, thus allowing for the generalization of most feasibility conclusions to patient/resident assessment in all four settings. With only a few noteworthy exceptions (i.e., observational assessments of cognitive function and mental status), there were very little missing data for any of the candidate SPADEs. It should be noted that these times are for the entire set of data elements (e.g., all four Nutritional Approaches, all eight PROMIS Anxiety data elements), so some SPADE sets took longer to complete partly because there were more data elements in the set. Although all candidate SPADEs took less than four minutes to complete, there was considerable variability, as follows:

- The Expression and Understanding, Hearing and Vision, and Nutritional Approaches SPADEs were each completed in *less than one minute* and were primarily perceived as having *very low assessment burden*, although the Nutritional Approaches data element set was perceived as having only moderately low burden.
- Assessment of the CAM, Behavioral Signs and Symptoms, Continence Interview, and Care Preferences (Interview and Chart) data element sets took *between one and two minutes* to complete and were perceived as having *low assessment burden*.
- The BIMS, PHQ-2 to 9, PROMIS Depression, PROMIS Anxiety, Special Treatments, Pain Interview, Staff Assessment of Pain, and Observational Assessment of Mental Status data element sets each required somewhere *between two and three minutes* to complete. Most of these were also perceived as having *high assessment burden*, except the Special Treatments and Staff Assessment of Pain were only moderately burdensome, and the BIMS and Pain Interview were reported to have low burden.
- The Continence Chart Review, PROMIS Global Health, Medication Reconciliation, and PHQ-9-OV all took *over three minutes* to complete and were perceived as having *high assessment burden*. However, the Continence Chart data element set was somewhat more

complicated to complete in the National Beta Test than it would be if implemented because of the look-back assessment component (Day 1, 3, 5, and 7).

Psychometric Performance

The psychometric performance of the candidate SPADEs varied but, for the most part, tended to be within acceptable ranges and in line with expectations.

Reliability

Interrater reliability was assessed based on ratings from paired assessors with kappa and weighted kappa coefficients, as well as with percent agreement. Kappas and weighted kappas were primarily moderate to excellent using standard interpretation guidelines (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).

All the candidate SPADEs that used patient interview for assessment had excellent kappa results (i.e., BIMS, PHQ-2 to 9, PROMIS Depression, PROMIS Anxiety, Pain Interview, Continence Interview, PROMIS Global Health, and Care Preferences Interview). This is not surprising given the paired assessor design employed. Specifically, for interview-based SPADEs, assessors conducted the assessment in pairs, with the facility/agency staff conducting the interview and both assessors (facility/agency staff and research nurse assessor) recording the patient's/resident's answer. The Observational Assessment SPADEs (Staff Assessment of Mental Status, PHQ-9-OV, and Observational Assessment of Pain or Distress) also had near-perfect kappas. At first blush, this is unexpected, given the challenge of rating patients/residents who are unable to communicate. However, these SPADEs include the coding option of "unknown/unable to assess." Thus, assessors likely chose this code whenever they were uncertain, leaving more concordance for the cases in which there was more confidence in the rating.

There were four candidate SPADE sets with all kappa coefficients in the substantial/good range (0.60–0.80), including the CAM, Nutritional Approaches, Hearing, and Continence Chart, and another group of candidate SPADE sets with kappas indicating moderate performance (0.41–0.60), including the two-data element version of Expression and Understanding, Vision, and Care Preferences Chart.

The remaining SPADE sets had kappas falling across the range of interpretative categories, as follows: kappas for the three-data element version of Expression and Understanding ranged from moderate to good (although the moderate kappa was 0.59, close to the lower bound for good), the Special Treatments data element kappas included one indicating poor performance (0.13) and the rest ranging from moderate to excellent, and the Medication Reconciliation SPADE set also had kappas ranging from moderate to excellent. With respect to these wide ranges of performance, the fact that excellent reliability can be achieved for many of the data elements within Special Treatments and Medication Reconciliation SPADE sets may indicate

that more-consistent training and guidance across the numerous data elements within the sets could improve overall reliability.

The percent agreement was calculated as a companion metric, and it varied fairly systematically according to the values of the kappas and weighted kappas. This concordance in the patterns of results lends strong credibility to the use of percent agreement alone to indicate reliability in cases where prevalence rates did not allow for calculation of kappa.

Validity

There are numerous indicators of the validity of the candidate SPADEs in the National Beta Test:

- The frequencies of responses and mean scores of each of the data elements across all assessment categories were very much in line with what would be expected for this sample of patients/residents receiving care in one of the four PAC settings. For example, rates of BIMS impairment were very much in line with expectations, considering the exclusion of patients/residents who were unable to communicate. Similarly, rates of positive screens on the PHQ-2 were as expected, as were rates of pain presence. In fact, the frequency distributions for all candidate SPADEs provide strong evidence for the validity of the data collection process.
- Evaluation of the associations of candidate SPADEs with patient/resident demographic and clinical characteristics, or known groups validity, also aligned fairly closely with expectations.
- On the whole, SPADE performance using different look-back periods in both the chart review and repeat assessment sub-study assessments indicated that the SPADEs are valid for use across the look-back periods tested. That is, for the chart review SPADEs, if something was present (e.g., oxygen therapy), it was noted in the chart on Day 1 for the majority of cases and extending the look-back period did not add additional information. For the SPADEs in the repeat assessment design, there was also little difference in frequency distributions across the three assessment points.
- Comparison of response frequencies from admission to discharge among patients who were assessed at both points indicated very little change in the candidate data elements tested. However, among patients/residents whose response did change, the changes were more likely to reflect improvement than decline. These results are in line with expectations.

Potential for Improving Quality

Evidence for clinical utility can be gleaned primarily from the assessor feedback, which varied considerably. Several candidate SPADE sets were rated by assessors as having *high clinical utility*, including the BIMS, Expression and Understanding, Behavioral Signs and Symptoms, Pain Interview, Hearing and Vision, and Continence Chart.

A second group was rated as having *moderate to high clinical utility*: PHQ-2 to 9, Care Preferences Interview, Care Preferences Chart, Medication Reconciliation, and Observational Assessment of Pain or Distress.

A third group was perceived as having *moderate clinical utility*, including the CAM, PROMIS Depression, PROMIS Anxiety, Nutritional Approaches, Special Treatments, and Continence Interview.

Finally, three data element sets were perceived as having only *somewhat to moderate clinical utility*. This group included PROMIS Global Health, Staff Assessment of Mental Status, and PHQ-9-OV.

All the candidate SPADEs that were selected for inclusion in the National Beta Test were evaluated by stakeholders and TEP members as having the potential to improve quality in one or more respects. Feedback from the assessors largely supported this sentiment. For example, many candidate SPADEs were noted as being particularly useful during care transitions (e.g., Expression and Understanding, Behavioral Signs and Symptoms, Medication Reconciliation), whereas others appeared to have potential for contributing to improvements in person-centered care and care planning (e.g., Continence Interview and Chart, Care Preferences, PHQ-2 to 9, PROMIS Depression and Anxiety).

Given the strong psychometric performance of the majority of candidate SPADEs, it is likely that many hold the potential to be used for quality comparisons in future work, should they be implemented.

Utility for Describing Case Mix

The results of the National Beta Test provide only minimal information regarding the utility of the candidate SPADEs for describing case mix. However, all the candidate SPADEs fall into one of the clinical categories specified by the IMPACT Act. These clinical categories were delineated in the IMPACT Act because they are potentially useful for contributing to the understanding of risk for patient/resident outcomes (e.g., risk of readmission, morbidity). Furthermore, inherent in cross-setting standardization is the possibility of development of comparable case mix adjustment models across settings. These advancements will certainly enhance the consistent description of case mix among patients and residents receiving care in one or more PAC settings.

Table 4.1 summarizes the performance of the data elements.

Table 4.1 Summary of Performance of Data Elements Tested in the National Beta Test

Category	Data Element	Data Source	Time to Complete (minutes)	Interrater Reliability		Assessor Feedback
			Mean (SD)	Kappa	Percent Agreement	
Cognitive Function	BIMS	Patient interview	2.2 (1.2)	0.83–0.93	94–98%	High clinical utility, low burden
Cognitive Function	CAM	Multiple sources, observation and chart	1.4 (0.7)	0.66	91–96%	Moderate clinical utility, moderately low burden
Cognitive Function	Expression and Understanding	Multiple sources, observation and chart	3–data element: 0.8 (0.4) 2–data element: 0.7 (0.3)	0.59–0.64 0.42, 0.32	93–95% 89%, 86%	High clinical utility, very low burden
Cognitive Function	Behavioral Signs and Symptoms	Multiple sources, observation and chart	1.4 (0.8)	—	95–100%	Clinically useful, especially for transfers
Mental Status	PHQ-2 to 9	Patient interview	2.3 (1.5)	0.95–1.00	97–100%	Moderate to high clinical utility, high burden
Mental Status	PROMIS Depression*	Patient interview	2.2 (0.8)	0.97–1.00	98–99%	Moderate clinical utility, relatively high burden
Mental Status	PROMIS Anxiety*	Patient interview	2.2 (0.8)	0.97–0.99	98–99%	Moderate clinical utility, relatively high burden
Special Services, Treatments, and Interventions	Nutritional Approaches	Multiple sources, observation and chart	0.88 (0.5)	0.60–0.65	80–100%	Moderate clinical utility, moderately low burden
Special Services, Treatments, and Interventions	Special Treatments	Multiple sources, observation and chart	2.4 (1.3)	0.13; 0.46–0.90	88–100%	Moderate clinical utility, moderate burden
Pain	Pain interview**	Patient interview	2.6 (1.4)	0.96–0.98	96–99%	High clinical utility, low burden
Impairments	Hearing and Vision	Patient interview	0.6 (0.3)	H: 0.65 V: 0.56	H: 84% V: 83%	High clinical utility, low burden
Impairments	Continence Interview	Patient interview	1.4 (0.7)	0.96–0.98	98–99%	Moderate clinical utility, moderately low burden

Category	Data Element	Data Source	Time to Complete (minutes)	Interrater Reliability		Assessor Feedback
			Mean (SD)	Kappa	Percent Agreement	
Impairments	Continence Chart	Multiple sources, observation and chart	3.5 (1.8)	0.66–0.79	74–100%	High clinical utility, high burden
Other Clinical Categories	PROMIS Global Health	Patient interview	3.6 (1.6)	0.95–0.99	95–99%	Some clinical utility, moderate burden
Other Clinical Categories	Care Preferences Interview	Patient interview	1.5 (0.7)	0.96, 0.96	98%, 99%	Moderate to high clinical utility, low burden
Other Clinical Categories	Care Preferences Chart	Multiple sources, observation and chart	1.2 (0.6)	0.56	83%	Moderate to high clinical utility, moderate burden
Other Clinical Categories	Medication Reconciliation	Multiple sources, observation and chart	3.2 (1.9)	0.42–0.89	79–97%	Moderate to high clinical utility, high burden
Observational Assessments	Staff Assessment of Mental Status	Patient observation	2.6 (1.6)	0.74–0.94	93–98%	Somewhat to moderate clinical utility, moderate burden
Observational Assessments	PHQ-9-OV	Patient observation	3.5 (1.7)	0.92–0.98	96–99%	Somewhat to moderate clinical utility, high burden
Observational Assessments	Observational Assessment of Pain or Distress	Patient observation	2.4 (1.7)	0.81–0.90	89–98%	Moderate to high clinical utility, moderate burden

NOTES: Interpretation of 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. H = Hearing, V = Vision.

* Results for combined versions (past 3 days and past 7 days).

** Results for combined versions (past 3 days and past 5 days).

References

- CAM—See Sharon K. Inouye, Christopher H. van Dyck, Cathy A. Alessi, Sharyl Balkin, Alan P. Siegal, and Ralph I. Horwitz, “Clarifying Confusion: The Confusion Assessment Method: A New Method for Detection of Delirium,” *Annals of Internal Medicine*, Vol. 113, No. 12, December 1990, pp. 941–948.
- 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.
- , “IMPACT Act Standardized Assessment National Testing,” webpage, last modified December 11, 2017e. As of August 5, 2019:
<https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/IMPACT-Act-of-2014/-IMPACT-Act-Standardized-Assessment-National-Testing-.html>

- , *Development and Maintenance of Post-Acute Care Cross-Setting Standardized Assessment Data: Public Comment Summary Report 2*, Baltimore, Md., January 2018.
- , “Long-Term Care Hospital PPS,” webpage, last modified March 8, 2019a. As of August 5, 2019:
<https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/LongTermCareHospitalPPS/index.html>
- , “Home Health PPS,” webpage, last modified March 8, 2019b. As of July 15, 2019:
<https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HomeHealthPPS/index.html>
- , “Long-Term Care Hospital (LTCH) Quality Reporting (QRP),” webpage, last modified April 16, 2019c. As of August 5, 2019:
<https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/lrch-quality-reporting/>
- , “Home Health Quality Reporting Requirements: Statutory Authority for Use of the OASIS Data Item Set and Home Health Quality Reporting,” webpage, last modified June 7, 2019d. As of August 5, 2019:
<https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/Home-Health-Quality-Reporting-Requirements.html>
- , “Inpatient Rehabilitation Facilities (IRF) Quality Reporting Program (QRP),” webpage, last modified July 2, 2019e. As of August 5, 2019:
<https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/IRF-Quality-Reporting/index.html>
- , “SNF Quality Reporting Program (IMPACT Act 2014),” webpage, last modified July 17, 2019f. As of August 5, 2019:
<https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/Skilled-Nursing-Facility-Quality-Reporting-Program/SNF-Quality-Reporting-Program-Overview.html>
- , “Inpatient Rehabilitation Facility PPS,” webpage, last modified July 31, 2019g. As of August 5, 2019:
<https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/InpatientRehabFacPPS/index.html>
- , “Skilled Nursing Facility PPS,” webpage, last modified August 1, 2019h. As of August 5, 2019:
<https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/SNFPPS/index.html>

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

- 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.
- 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, March 2017. As of February 6, 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 2018a. As of February 6, 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, Stacy Mandel, Emily Chen, et al., CMS/RAND stakeholder forum transcript, November 2018b. As of August 12, 2019:
https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/TTtranscript-of-11_27_18-Forum.pdf
- Edelen, Maria Orlando, Anthony Rodriguez, Sangeeta C. Ahluwalia, Emily K. Chen, Catherine C. Cohen, Sarah Dalton, Jason Michel Etchegaray, Wenjing Huang, Brenda Karkos, David J. Klein, Jaime Madrigano, Monique Martineau, Terry Moore, Patrick Orr, Jessica Phillips, Ben Schalet, Victoria Shier, Susan M. Paddock, and Denis Agniel, *Development and Evaluation of Candidate Standardized Patient Assessment Data Elements: Findings from the National Beta Test (Volume 2: Background and Methods)*, Santa Monica, Calif.: RAND Corporation, RR-3004-CMS, 2019a. As of August 20, 2019:
https://www.rand.org/pubs/research_reports/RR3004.html
- , *Development and Evaluation of Candidate Standardized Patient Assessment Data Elements: Findings from the National Beta Test (Volume 4: Cognitive Function)*, Santa

Monica, Calif.: RAND Corporation, RR-3004/3-CMS, 2019b. As of August 20, 2019:
https://www.rand.org/pubs/research_reports/RR3004z3.html

———, *Development and Evaluation of Candidate Standardized Patient Assessment Data Elements: Findings from the National Beta Test (Volume 5: Mental Status and Pain)*, Santa Monica, Calif.: RAND Corporation, RR-3004/4-CMS, 2019c. As of August 20, 2019:
https://www.rand.org/pubs/research_reports/RR3004z4.html

———, *Development and Evaluation of Candidate Standardized Patient Assessment Data Elements: Findings from the National Beta Test (Volume 6: Impairments and Special Services, Treatments, and Interventions)*, Santa Monica, Calif.: RAND Corporation, RR-3004/5-CMS, 2019d. As of August 20, 2019:
https://www.rand.org/pubs/research_reports/RR3004z5.html

———, *Development and Evaluation of Candidate Standardized Patient Assessment Data Elements: Findings from the National Beta Test (Volume 7: Global Health, Care Preferences, and Medication Reconciliation)*, Santa Monica, Calif.: RAND Corporation, RR-3004/6-CMS, 2019e. As of August 20, 2019:
https://www.rand.org/pubs/research_reports/RR3004z6.html

———, *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, 2019f. As of August 20, 2019:
https://www.rand.org/pubs/research_reports/RR3004z7.html

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

Medicare Payment Advisory Commission, *Report to the Congress: Medicare and the Health Care Delivery System*, Washington, D.C., June 2018. As of July 15, 2019:
http://medpac.gov/docs/default-source/reports/jun18_medpacreporttocongress_sec.pdf

———, *Report to the Congress: Medicare Payment Policy*, Washington, D.C., March 2019.

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.

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.

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 February 6, 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 March 11, 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>

———, “Event to Share Findings from RAND’s IMPACT National Beta Test (Standardized Data Elements; November 27, 2018): Q&A,” December 18, 2018. As of August 12, 2019:
https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/QA_RAND-Forum-on-PAC-SPADES.pdf

———, “Input on Standardized Patient Assessment Data Elements (SPADES) Received After November 27, 2018 Stakeholder Meeting,” February 1, 2019. As of August 12, 2019:
<https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/Input-on-SPADES-Received-After-November-27-2018-Stakeholder-Meeting.pdf>

RAND Health Care, “Early Findings from the RAND IMPACT National Beta Test of Candidate Standardized Patient Assessment Data Elements (SPADES),” November 27, 2018. As of August 12, 2019:
https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/Forum-Slides_Early-Findings-from-the-RAND-IMPACT-National-Beta-Test.pdf

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.