

3-2012

Post-Acute Care Payment Reform Demonstration: Final Report Volume 2 of 4

Barbara Gage

Melissa Morley


Laura Smith

Melvin Ingber

Anne Deutsch

See next page for additional authors

Follow this and additional works at: https://hsrc.himmelfarb.gwu.edu/smhs_crl_facpubs

 Part of the [Health and Medical Administration Commons](#), [Health Services Administration Commons](#), and the [Health Services Research Commons](#)

APA Citation

Gage, B., Morley, M., Smith, L., Ingber, M., Deutsch, A., Kline, T., Dever, J., Abbate, J., Miller, R., Lyda-McDonald, B., Kelleher, C., Garfinkel, D., Manning, J., Murtaugh, C. M., Stineman, M., & Mallinson, T. (2012). Post-Acute Care Payment Reform Demonstration: Final Report Volume 2 of 4., (). Retrieved from https://hsrc.himmelfarb.gwu.edu/smhs_crl_facpubs/108

This Report is brought to you for free and open access by the Clinical Research and Leadership at Health Sciences Research Commons. It has been accepted for inclusion in Clinical Research and Leadership Faculty Publications by an authorized administrator of Health Sciences Research Commons. For more information, please contact hsrc@gwu.edu.

Authors

Barbara Gage, Melissa Morley, Laura Smith, Melvin Ingber, Anne Deutsch, Tracy Kline, Jill Dever, Judith Abbate, Richard Miller, Brienne Lyda-McDonald, Cynthia Kelleher, Danielle Garfinkel, Joshua Manning, Christopher M. Murtaugh, Margaret Stineman, and Trudy Mallinson

March 2012

**Post-Acute Care Payment Reform Demonstration:
Final Report
Volume 2 of 4**

Prepared for

Shannon Flood

Centers for Medicare & Medicaid Services
Center for Medicare & Medicaid Innovation
7205 Windsor Boulevard
Baltimore, MD 21244-1850

Prepared by

Barbara Gage, PhD

Melissa Morley, PhD

Laura Smith, PhD

Melvin J. Ingber, PhD

Anne Deutsch, RN, PhD, CRRN

Tracy Kline, PhD

Jill Dever, PhD

Judith Abbate, PhD

Richard Miller, PhD

Brieanne Lyda-McDonald, MSPH

Cynthia Kelleher, MBA

Danielle Garfinkel, BA

Joshua Manning, BS

RTI International

3040 Cornwallis Road

Research Triangle Park, NC 27709

Christopher M. Murtaugh, PhD

Visiting Nurse Service of New York

Margaret Stineman, MD

University of Pennsylvania

Trudy Mallinson, PhD, OTR/L, NZROT

University of Southern California

RTI Project Number 0209853.005.001

Post-Acute Care Payment Reform Demonstration Final Report

by Barbara Gage, PhD, Project Director
Melissa Morley, PhD
Laura Smith, PhD
Melvin J. Ingber, PhD
Anne Deutsch, RN, PhD, CRRN
Tracy Kline, PhD
Jill Dever, PhD
Judith Abbate, PhD
Richard Miller, PhD
Brieanne Lyda-McDonald, MSPH
Cynthia Kelleher, MBA
Danielle Garfinkel, BA
Joshua Manning, BS
RTI International
3040 Cornwallis Road
Research Triangle Park, NC 27709
Christopher M. Murtaugh, PhD
Visiting Nurse Service of New York
Margaret Stineman, MD
University of Pennsylvania
Trudy Mallinson, PhD, OTR/L, NZROT
University of Southern California

Federal Project Officer: Shannon Flood

CMS Contract No. HHSM-500-2005-00029I

March 2012

This project was funded by the Centers for Medicare & Medicaid Services under contract no. HHSM-500-2005-00029I. The statements contained in this report are solely those of the authors and do not necessarily reflect the views or policies of the Centers for Medicare & Medicaid Services. RTI assumes responsibility for the accuracy and completeness of the information contained in this report.

CONTENTS SUMMARY

This document represents Volume 2 of 4 of the final report for the Post-Acute Care Payment Reform Demonstration (PAC-PRD). This project was conducted by RTI International under contract with the Centers for Medicare & Medicaid Services. The report has 12 sections, which are divided into four volumes.

- Volume 1: Executive Summary
- Volume 2: Sections 1–4
 - Section 1: Introduction
 - Section 2: Underlying Issues of the PAC-PRD Initiating Legislation
 - Section 3: Developing Standardized Measurement Approaches: The Continuity Assessment Record and Evaluation (CARE)
 - Section 4: Demonstration Methods and Data Collection
- Volume 3: Sections 5–6
 - Section 5: Framework for Analysis
 - Section 6: Factors Associated With Hospital Discharge Destination
- Volume 4: Sections 7–12; References
 - Section 7: Outcomes: Hospital Readmissions
 - Section 8: Outcomes: Functional Status
 - Section 9: Determinants of Resource Intensity: Methods and Analytic Sample Description
 - Section 10: Determinants of Resource Intensity: Lessons From the CART Analysis
 - Section 11: Determinants of Resource Intensity: Multivariate Regression Results
 - Section 12: Conclusions and Review of Findings
 - References

This page intentionally left blank.

CONTENTS

Section 1 Introduction.....	1
Section 2 Underlying Issues of the PAC-PRD Initiating Legislation.....	7
2.1 Beneficiaries’ Use of Post-Acute Services	9
2.2 Ability to Compare Patients Across Settings.....	11
2.3 Inconsistencies in Case-Mix Systems and Unintended Incentives.....	13
2.3.1 Consistent Measurement of Items Included in Case-Mix Systems	15
2.3.2 Nonrepresented or Underrepresented Patient Severity Measures.....	15
2.3.3 Types of Patient Costs Modeled	15
2.3.4 Unit of Payment	15
2.3.5 Service Use Measures	19
2.3.6 Course of Treatment Perspectives	19
2.3.7 Incorporating Issues of Value Into Payment.....	19
2.4 Other Rules and Factors.....	20
Section 3 Developing Standardized Measurement Approaches: The Continuity Assessment Record and Evaluation (CARE).....	21
3.1 Introduction to the CARE Tool.....	21
3.1.1 Stakeholder Input.....	22
3.1.2 Item Selection	22
3.1.3 Response Rate.....	22
3.2 Guiding Principles of CARE Tool Development	23
3.3 Development of the CARE Tool and CARE Tool Items.....	26
3.3.1 Defining the Domains	26
3.3.2 Forming Clinical Workgroups.....	27
3.3.3 Selecting Items for Use in the CARE Tool.....	28
3.3.4 Basic Organization of the CARE Tools.....	29
3.4 CARE Item Description.....	30
3.4.1 Administrative Items.....	30
3.4.2 Premorbidity Patient Information	30
3.4.3 Current Medical Information	30
3.4.4 Interview Items: Cognitive Status, Mood, and Pain	31
3.4.5 Impairments	32
3.4.6 Functional Status.....	32
3.4.7 Frailty and Life Expectancy.....	33
3.4.8 Discharge Information	34
3.4.9 ICD-9 Codes	34
3.5 Testing and Feedback During Development.....	34

3.5.1	Pilot Tests.....	35
3.5.2	Stakeholder and Public Comment During CARE Development.....	35
3.5.3	Stakeholder and Public Comment During CARE Use	38
3.6	Reliability Study	38
3.6.1	Traditional Inter-Rater Reliability Testing	39
3.6.2	Results.....	41
3.7	Reliability Testing of Clinician Agreement Across Settings.....	45
3.7.1	Prior Functioning.....	46
3.7.2	Skin Integrity	46
3.7.3	Cognitive Status, Mood, and Pain.....	46
3.7.4	Impairments	46
3.7.5	Functional Status.....	47
3.8	Summary of the Results.....	48
3.9	Next Steps: Use of a Flexible Electronic Standards-Based Instrument.....	48
3.10	Summary of the Section.....	49
Section 4 Demonstration Methods and Data Collection.....		51
4.1	Sample Framework.....	51
4.1.1	Market Area Characteristics	51
4.1.2	Market Area Selection	52
4.2	PAC Utilization and Episodes of Care in Initial PAC-PRD Market Areas	54
4.2.1	Use of PAC Services and Use of Specific Provider Types in Participating Markets.....	54
4.2.2	Trajectory of PAC Service Use in Participating Initial Collection Markets	55
4.3	Provider Selection and Recruitment	56
4.4	Data Collection	57
4.4.1	CARE Data Collection.....	57
4.4.2	CRU Data Collection.....	59

List of Figures

Figure 2-1	Post-acute care transitions after acute hospital discharge, 2008.....	10
Figure 4-1	Nationwide map of initial collection markets and supplemental collection cities...	62

List of Tables

Table 2-1	Upper body dressing functional scales in current assessment instruments.....	12
Table 2-2	Selected advantages and disadvantages of payment unit alternatives	16
Table 3-1	IRR testing providers by type/level of care	39
Table 3-2	Standard interpretation of kappa scores.....	40
Table 3-3	Video reliability testing providers by type/level of care.....	45
Table 4-1	Count of providers, by provider type, by initial collection market, 2006.....	63

Table 4-2	Percentage of beneficiaries discharged to post-acute care and first site of care, by market area, 2006.....	64
Table 4-3	Market area episode descriptives, 2006.....	65
Table 4-4a	Market area episode descriptives, by service type (HHA, IRF, and SNF), 2006	66
Table 4-4b	Market area episode descriptives, by service type (LTCH, HOPD, and readmission), 2006.....	67
Table 4-5a	Top 10 episode patterns in the initial PAC-PRD sample for beneficiaries initiating service use in an acute hospital	68
Table 4-5b	Top five episode patterns in the initial PAC-PRD sample for beneficiaries initiating service use in an HHA.....	69
Table 4-6	Count of participating providers, by provider type, by market area.....	70
Table 4-7a	Initial sample CARE assessment counts by assessment type, by provider type.....	71
Table 4-7b	Supplemental providers CARE assessment counts by assessment type, by provider type	71
Table 4-8	Number of CRU patients, patient-days, and days per patient, by setting	72

This page intentionally left blank.

SECTION 1 INTRODUCTION

This report is the final report for the Post-Acute Care Payment Reform Demonstration (PAC-PRD). This project was conducted by RTI International under contract with the Centers for Medicare and Medicaid Services. This report builds on the May 2011 PAC-PRD Report to Congress (RTC) (http://www.cms.gov/reports/downloads/Flood_PACPRD_RTC_CMS_Report_Jan_2012.pdf) and the associated supplemental report (http://www.cms.gov/Reports/Downloads/GAGE_PACPRD_RTC_Supp_Materials_May_2011.pdf). This report includes four new chapters examining discharge destinations and refining the resource intensity analysis. Data from the second phase of the data collection are also incorporated in this report.

The PAC-PRD was authorized in the Deficit Reduction Act (DRA) of 2005 (S. 1932, Title V, Sec. 5008) to provide Congress information on Medicare program costs, patient outcomes, and other factors associated with treatment in different post-acute care (PAC) sites. The DRA called for a standardized patient assessment instrument to be used at discharge from the hospital and admission to and discharge from PAC sites. Standardizing the current assessment items was necessary to compare patients' clinical characteristics across settings and to examine the factors associated with costs and resource use, outcomes, discharge placement, and good care transitions across an episode of care. Information on both the fixed and variable costs associated with caring for patients in each site was also of interest. The results are intended to provide information on ways to improve the consistency of payment incentives across a Medicare beneficiary's episode of care.

Participating providers included five settings: acute care hospitals and the four types of PAC providers covered under Medicare Part A insurance: long-term care hospitals (LTCHs), inpatient rehabilitation facilities (IRFs), skilled nursing facilities (SNFs), and home health agencies (HHAs).

Work on the PAC-PRD included three major components.

1. To meet the Congressional mandate, a single, comprehensive patient assessment instrument needed to be developed. In Component One, a select set of standardized items commonly used at intake assessment in each setting was identified through input from the five clinical communities (acute hospitals and the four types of PAC sites), including numerous technical expert panels, pilot tests, and a large-scale national demonstration. This set formed the basis of the Continuity Assessment Record and Evaluation (CARE) items.
2. Component Two developed a secure, Internet-based software application for collecting CARE data from the participating providers. This tool allowed patient information to be shared across sites, potentially allowing the admitting clinician to view the assessment completed at prior settings treating the patient. Effectively, this component provided the infrastructure for a virtual electronic health record for Medicare beneficiaries participating in the PAC-PRD.

3. Component Three collected data in the five settings of interest and performed the associated analysis of the demonstration. Almost 54,000 assessments were collected from 190 providers over the course of the entire data collection period.¹ Data were tested for reliability and used in payment and outcome analysis to understand differences in patients treated in each of the four PAC sites, the resources associated with those treatments, and the outcomes resulting from each service.

This technical report provides in-depth analysis of the findings submitted by the Centers for Medicare & Medicaid Services in its RTC as well as additional information on factors associated with each hospital discharge destination and a refined analysis of the resource intensity results provided in the May 2011 report. The report has 12 sections, which are divided into four volumes. Volume 1 is the Executive Summary; Volume 2, Sections 1–4; Volume 3, Sections 5–6; and Volume 4, Sections 7–12 and the references for the whole report. A detailed description of the sections follows.

- **Section 1: Introduction.** Provides an overview of the 12 sections of the report.
- **Section 2: Underlying Issues of the PAC-PRD Initiating Legislation.** Discusses the issues identified in the initiating legislation leading to the need for PAC-PRD, including
 - the need for a standardized measure of patient acuity and resource use across PAC settings to examine such issues as the differences in populations treated in the four PAC settings, the resources provided, the outcomes gained, and whether the Medicare payment rates and beneficiary outcomes differed when similar patients were treated in more than one setting; and
 - variations in the current payment systems that may lead to unwarranted inconsistencies between the various PAC settings.
- **Section 3: Developing Standardized Measurement Approaches: The Continuity Assessment Record and Evaluation (CARE).** Presents the measurement approach for developing standardized assessment items to measure beneficiaries' medical, functional, and cognitive status, including information on
 - development of the standardized measurement approach, including development and pilot testing of the CARE items and assessment time frames in the five settings²; inclusion of stakeholder input throughout the process; final item selection and development and their relationships to items currently used in hospital assessments; and the three Federal assessment tools: the Minimum Data

¹ The RTC was based on data from the first 140 providers. Additional sites were admitted to the demonstration in the supplemental phase of data collection.

² Many of the PAC prospective payment systems (PPSs) included similar concepts, but the specific item varied by PPS. Almost all of the CARE items are commonly assessed in all five settings during an admission process.

Set (MDS), Outcome and Assessment Information Set (OASIS), and Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI); and

- testing of the CARE items for validity and reliability, including two test approaches and the results, which show how the standardized items performed across settings, both within and across clinical sites.
- **Section 4: Demonstration Methods and Data Collection.** Updates the May sample description to provide information on participating organizations in both the initial and supplemental phases of data collection. This chapter provides an overview of the primary data collection approach, including
 - sample selection process—how markets and providers were selected;
 - data collection process—methodology for collecting the CARE and cost and resource utilization (CRU) staff time study data; and
 - representativeness of the PAC-PRD sample relative to PAC users nationally, including market-level comparisons of episode patterns, spending, and utilization.
- **Section 5: Framework for Analysis.** Provides the conceptual framework for understanding the analytic approach, building on the existing approaches for defining patient complexity to explain variation in cost and outcomes, including
 - development of a case-mix classification framework and discussion of the three classification domains (medical, functional, and cognitive) and how these factors are used in the current PAC prospective payment systems (PPSs) to classify patient complexity; and
 - definition of more complex concepts and discussion of how to operationalize them, including the primary reason for treatment, comorbidities, and functional status, including refinements to some of the measures since the Interim Report.
- **Section 6: Factors Associated With Hospital Discharge Destination.** Presents sample descriptions of the cases directly discharged from acute care to each type of PAC setting and multivariate analysis of similarities and differences across sites of care. Analyses examine the total post-hospital sample and several subgroups of populations, including those with neurological disorders, respiratory conditions, circulatory conditions, and musculoskeletal conditions. Several analyses are presented on each, including
 - any PAC use: examines case-mix differences between those who go home without subsequent Medicare services (excluding physician services) and those who transfer to LTCHs, IRFs, SNFs, or HHAs;
 - SNF or IRF use: examines case-mix differences between the two types of patients discharged directly from acute care to these services;

- SNF or HHA use: examines case-mix differences between the two types of patients discharged directly from acute care to these services; and
- discharge to LTCH, IRF, SNF, or HHA relative to no subsequent Medicare service (excluding physician services): examines similarities and differences in patient populations across the service system.
- **Section 7: Outcomes: Hospital Readmissions.** Presents findings related to whether medical outcomes, such as the probability of readmission, are related to the PAC setting, after controlling for patient characteristics.
- **Section 8: Outcomes: Functional Status.** Uses the standardized function items to examine functional impairment levels at admission and discharge in each PAC setting and to examine whether functional status outcomes differ by PAC setting after controlling for patient acuity (medical, functional, and cognitive). Two types of functional outcomes are examined:
 - Self-care status is based on items measuring abilities in eating, oral hygiene, toilet hygiene, dressing upper body, dressing lower body, putting on and taking off footwear, washing upper body, and showering or bathing oneself.
 - Mobility status is based on items measuring abilities in moving from lying to sitting on side of bed, from sitting to standing, from chair or bed to chair, and from sitting to lying; moving to a toilet or to a car; rolling from left to right; picking up objects; and taking steps (1, 4, or 12 steps), as well as walking or wheelchair mobility at different distances and on uneven surfaces.
- **Section 9: Determinants of Resource Intensity: Methods and Analytic Sample Description.** Describes resource intensity measure development and provides description of the resource intensity sample. In addition, this chapter reviews the resource intensity results presented in the RTC. Two measures of resource intensity are used:
 - routine resource intensity, including all nontherapy staff whose costs are embedded in general per-diem costs (e.g., nursing, case management, respiratory therapy, aides); and
 - therapy resource intensity, including the licensed and registered physical therapy, occupational therapy, and speech pathology clinicians.
- **Section 10: Determinants of Resource Intensity: Lessons From the CART analysis.** Building on the May 2011 analysis, describes exploratory work that uses a classification and regression tree (CART) approach to create subsamples with similar resource intensity and to explore refinements of the independent variables used in the earlier regression analysis. The results of this section are incorporated into the regression models presented in Section 11.

- **Section 11: Determinants of Resource Intensity: Multivariate Regression Results.** This section applies the results of the CART analysis to refine the multivariate models discussed in the May 2011 report. Separate models are developed for routine and therapy intensity.
- **Section 12: Conclusions and Review of Findings.** Reviews the findings and discusses the conclusions associated with the analyses as they relate to the Federal initiative to create more consistent incentives, measurement approaches, and payment policies.

This page intentionally left blank.

SECTION 2

UNDERLYING ISSUES OF THE PAC-PRD INITIATING LEGISLATION

Almost one in five Medicare beneficiaries is admitted to the hospital each year; among them almost 35 percent will be discharged from the hospital to one of four post-acute care (PAC) sites for additional nursing or therapy treatments. These PAC sites include long-term care hospitals (LTCHs), inpatient rehabilitation facilities (IRFs), skilled nursing facilities (SNFs), and home health agencies (HHAs). Many patients may continue on to additional PAC sites after the first service. In general, the four PAC sites are assumed to differ in the type and intensity of services provided, effectively providing a continuum of care. But these providers' services are not mutually exclusive: each of the three inpatient PAC settings (LTCHs, IRFs, and SNFs) provides 24-hour nursing, and all four settings provide physical, occupational, and speech pathology services to some extent.

Although the four PAC sites each have different Medicare coverage rules, minimum certification standards, and prospective payment systems (PPSs), they overlap in providing nursing and therapy services to the Medicare population. Past research has shown that patients treated for the same condition in an acute hospital may be discharged to different types of PAC for subsequent treatment, depending on the availability of PAC options in the local market and other factors not measurable in the Medicare claims data (Gage, 1999; Gage et al., 2009).

Three of the PAC PPSs (IRF, SNF, and HHA) are based on assessment data that measure patient complexity factors not found in the claims data. Although the concepts are similar in each PPS, the exact items used to measure patient complexity differ across the three systems. The fourth PPS (LTCH) relies entirely on claims data for measuring severity, limiting measures largely to diagnoses and procedures data.

Because each PAC PPS uses different case-mix measurement items, it has been difficult to compare the populations admitted to each site and the costs and outcomes associated with treatment in the four PAC sites. These issues are further complicated by the different episode patterns, which may include several types of PAC service use during an episode of care and, depending on local availability, may use alternative types of settings for similar services. These issues underlie the need to ensure appropriate payment incentives for each of the PAC providers, because they may be used together as part of a beneficiary's complete episode of care.

The structure of Medicare PAC payment policies has fundamentally changed over the past decade. In particular, all PAC providers moved from cost-based reimbursement toward more bundled payment systems, such as the PPSs currently in place. The potential benefits of bundled systems were highlighted through experiences in acute inpatient settings, where providers received financial reimbursement for diagnosis-related groups (DRGs) based on inpatient episodes of care. After the Balanced Budget Act (BBA) of 1997, PAC providers moved to similar bundled payment systems (Cotterill and Gage, 2002). SNFs were the first to move to a case-mix-adjusted PPS in 1998, followed by HHAs in 2000 and by IRFs and LTCHs in 2002.

Under the PPSs, providers are encouraged to manage resources and simultaneously achieve desired outcomes. Each PPS evolved separately and uses site-specific case-mix

adjustment systems and case-mix data collection tools (Gage and Green, 2006). The resulting absence of a standardized case-mix measurement system that could be applied in all medical and functional rehabilitation settings has restricted the Medicare program's ability to consider the effects of an episode of care when paying for services. Instead, much of the research is site specific and uses the case-mix tools that are designed for each system to examine costs and quality within each site of care.

Several studies have examined the impact of changing payment policies, either across providers (Gage, Morley, and Green, 2007; Gage, 1999; Gage, Bartosch, and Osber, 2005; Beeuwkes Buntin et al., 2005) or for specific providers (Liu and Black, 2003; White, 2003; Pizer, White, and White, 2002; McCall et al., 2003; Medicare Payment Advisory Commission [MedPAC], 2004). Some have focused on the effects of the home health payment changes (McCall et al., 2003; Gage, 1998; Zhu et al., 2004; Murtaugh et al., 2003), changes related to the effects of the IRF PPS (Beeuwkes Buntin et al., 2005; Gage et al., ongoing), or SNF payment policy changes (Liu and Black, 2003; Stearns, Dalton, and Holmes, 2006; Gilman, Gage, and Osber, 2005). The LTCH PPS, which is the newest system, has undergone the fewest post-PPS studies (Dalton and Gage, 2007; MedPAC, 2004; Gage, Pilkauskas et al., 2005).

A more recent study by Gage et al. (2007) constructed episodes of care to examine post-acute patterns and first-site-of-care decisions, given the complex incentives under PPSs for HHAs, IRFs, LTCHs, and SNFs in 2005. The study examined discharge patterns for different types of hospital cases. Two severity measures were used: (1) the All Patient Refined DRGs (APR-DRGs) developed by 3M Health Information Systems for inpatient hospital quality and mortality studies, and (2) the Hierarchical Condition Categories (HCCs) developed for payment of Medicare managed care organizations. The measures are based on principal and secondary diagnosis or prior service use, respectively. Together these severity measures capture resource utilization associated with inpatient hospital stay (APR-DRG), the degree to which future (next year's) utilization of health care is predicted by prior service use, or the effects of chronic health conditions (HCCs) and differences in medical acuity. This study revealed significant differences in populations using each type of PAC service. Although the services provided by each setting may be similar, certain factors distinguished differences in the probability of using each type of service. These included differences in diagnoses, sociodemographic factors, severity measures, PAC supply, and regional location.

Adequately controlling for case-mix severity is key to understanding the differences in populations receiving PAC services and the appropriateness of the incentives in each of the four PPSs. The payment and coverage policies clearly distinguish between certain patients' treatment needs, but they are less distinctive for a substantial number of hospital discharges who may be treated in multiple settings, depending, in part, on the types of services provided by individual IRFs, SNFs, and LTCHs. Also, these similarities in the types of services provided in these inpatient settings raise concern that PAC providers may be providing substitute services while receiving substantially different payments for those services (MedPAC, 2004; Gage et al., 2005). Furthermore, despite these similarities, the certification requirements that protect beneficiary quality of care differ by provider, suggesting that although two similar patients may be treated in differently licensed providers, the required licensure standards may be different, as may the costs of care and outcomes.

Although the argument has been made that patients treated in these different settings vary in terms of their acuity, little empirical evidence exists to support the hypothesis. The absence of consistent severity measures in the PAC assessment tools has contributed to the difficulties in examining severity as it relates to site-of-care choices, treatment intensity, and outcomes (Gage and Green, 2006).

2.1 Beneficiaries' Use of Post-Acute Services

Data collection for the PAC-PRD began in 2008. This section provides background information on the use of PAC services in 2008 in order to provide context for the patterns of utilization seen in this demonstration. In 2008, 38 percent of Medicare beneficiaries discharged from a general acute care hospital continued into a PAC site. Of the patients receiving PAC services, 37.4 percent were discharged to an HHA, 42.2 percent to an SNF, 8.6 percent to an IRF, and 1.7 percent to an LTCH. The remaining patients received therapy services in either a hospital outpatient department or a therapist's office (**Figure 2-1**).

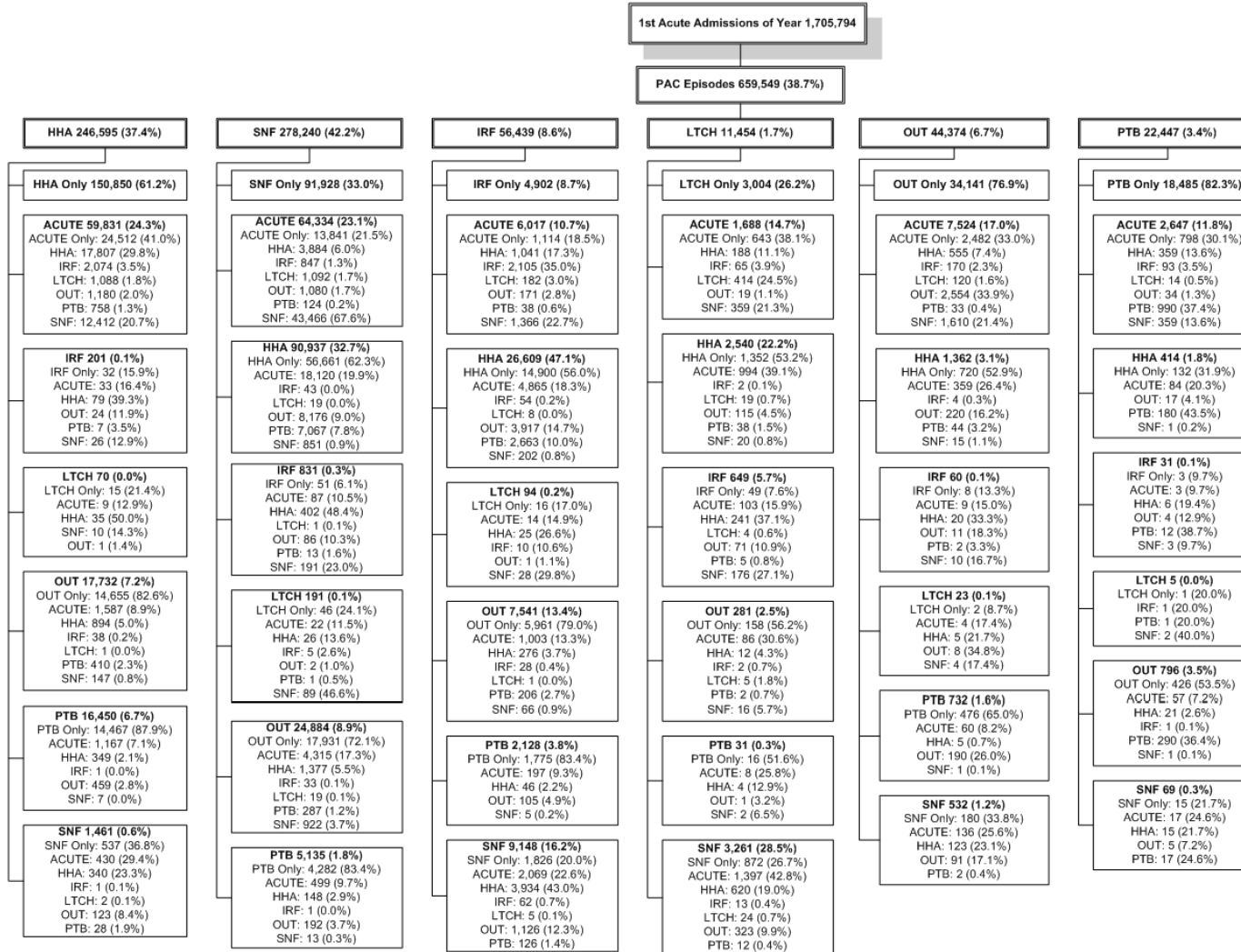
A large number of those discharged to PAC used more than one service during their episode of care, particularly those discharged to SNFs and LTCHs. For example, 67 percent of those discharged to SNFs continued on to additional services. Almost a quarter of them were readmitted to the acute hospital (23.1 percent). Another third (32.7 percent) were discharged from the SNF to an HHA. In patients with the Acute-SNF-HHA pattern, almost 20 percent (19.9 percent) returned to the acute hospital within 30 days of discharge from the HHA.

LTCH patients were also likely to use multiple types of PAC services. About 74 percent of cases discharged to LTCHs were discharged to additional services after leaving the LTCH, either back to the acute hospital (14.7 percent) or on to an HHA (22.2 percent), IRF (5.7 percent), or SNF (28.5 percent). A substantial share of each of the cases discharged from an LTCH to a third PAC setting were readmitted to the hospital within 30 days of discharge from the PAC service, ranging from 15.9 percent (LTCH-to-IRF cases) to 42.8 percent (LTCH to SNF).

Hospital patients discharged to IRFs were also likely to use multiple PAC services, although the most common third sites of care were in the community. Almost half of the acute-to-IRF cases (47.1 percent) were discharged from the IRF to an HHA; another 17.2 percent were discharged to outpatient or independent therapy. About 16.2 percent were discharged from the IRF to an SNF, and less than 1 percent of these returned to the IRF.

Acute-to-HHA cases typically used only the one service (61.2 percent) unless they were readmitted to the acute hospital (24.3 percent). Of the readmitted cases, 29.8 percent were readmitted to the HHA, and 20.7 percent were discharged instead to an SNF. In examining the home health patterns, it is important to keep in mind that a significant number of the home health population does not come through an acute admission or as part of a post-acute trajectory of care but instead are directly admitted to the HHA from the community. Similarly, those discharged from the hospital to outpatient therapy (6.7 percent) or other independent therapists (3.4 percent) typically used only that one post-hospital service.

Figure 2-1
Post-acute care transitions after acute hospital discharge, 2008



10

NOTE: The sample includes 1,705,794 beneficiaries with index acute hospitalizations in 2008 (30 percent of all index acute hospitalizations). An index acute hospitalization is defined as an acute hospitalization following a 30-day period without acute, skilled nursing facility (SNF), inpatient rehabilitation facility (IRF), long-term care hospital (LTCH), or home health agency (HHA) service use. Acute, SNF, IRF, LTCH, HHA, and therapy service use, both hospital outpatient department (OUT) and independent therapists (PTBs), were followed under a 30-day variable-length episode definition, which included all services before a 30-day gap in service use. The number and percentage of beneficiaries in each trajectory are shown here.

SOURCE: RTI analysis of 2008 Medicare claims (random 30 percent sample of acute initiating events).

2.2 Ability to Compare Patients Across Settings

An important issue in this demonstration is the potential variation in payment levels across provider settings for the same type of patient. To examine variation in payment levels, one must be able to assess and risk-adjust patient severity and needs in a consistent manner across settings.

The Medicare program currently mandates that IRFs, nursing facilities (including SNFs), and HHAs each submit assessment data on the beneficiary's medical, functional, and cognitive status. These mandated, site-specific patient assessment tools are referred to as the IRF-PAI (Patient Assessment Instrument), MDS (Minimum Data Set), and OASIS (Outcome and Assessment Information Set), respectively. These instruments are used to measure patient severity at admission and during different times in the patient treatment. The information collected through these assessments is used by the Centers for Medicare & Medicaid Services (CMS) to calculate payment groups, generate quality measures, and monitor regulatory compliance. Additionally, many states use data from these assessments for Medicaid payment and quality monitoring purposes.

Although the three mandated assessments measure similar concepts, they use different clinical items, different assessment time frames, and disparate measurement scales to assess health, physical function, and cognitive status. Acute care hospitals and LTCHs collect similar information at intake but are not currently required to submit these data to CMS. Instead, payments from acute care hospitals and LTCHs are based on claims information, which identifies the precipitating acute event and associated procedures.

Among the three mandated assessment tools, the variation in measurement techniques makes it difficult to compare patients treated in these different settings. For example, the different ways that function items are measured has complicated the ability to compare patients assessed in the different settings. The three instruments vary in whether they assess the patient's best performance or worst performance. In addition, the observation windows vary from assessing performance in activities of daily living (ADLs) over the past 7 days to the current day only. The measures differ in the number of ADL activities assessed (from 8 to 18) and the exact definitions of the activity. Finally, the three instruments vary in the scales used to assess performance. The different scales are illustrated in **Table 2-1**. The use of uniform items can standardize these case-mix measurement approaches and allow empirical consideration of differences in complexity of patients treated in different settings. The use of a standardized assessment tool in acute hospitals and PAC settings will allow for the comparison of functional outcomes across settings, for the tracking of outcomes from the beginning of a trajectory of care to final discharge, and for the improved communication of patient information between settings at the time of transfer.

Table 2-1
Upper body dressing functional scales in current assessment instruments

IRF-PAI	MDS 3.0	OASIS-C
<p>“Dressing—Upper Body” includes dressing and undressing above the waist, as well as applying and removing a prosthesis or orthotic when applicable. The patient performs this activity safely.</p> <p>7 = Complete independence (timely, safely)</p> <p>6 = Modified independence (with device)</p> <p>5 = Supervision (subject does 100% but with supervision)</p> <p>4 = Minimal assistance (subject does 75% or more)</p> <p>3 = Moderate assistance (subject does 50%–75%)</p> <p>2 = Maximal assistance (subject does 25%–49%)</p> <p>1 = Total assistance (subject does less than 25%)</p> <p>0 = Activity does not occur</p>	<p>“Dressing”: how resident puts on, fastens, and takes off all items of street clothing, including donning/removing a prosthesis.</p> <p>Assess including the resident’s performance when using adaptive devices.</p> <p>0 = Independent</p> <p>Measure does not separate out use of assistive devices.</p> <p>1 = Supervision (oversight, encouragement, or cueing)</p> <p>2 = Limited assistance (guided maneuvering)</p> <p>3 = Extensive assistance (weight-bearing assistance or total assistance some but not all of the time)</p> <p>4 = Total dependence</p> <p>7 = Activity occurred only once or twice in observation window</p> <p>8 = Activity did not occur</p>	<p>“Current Ability to Dress Upper Body Safely” (with or without dressing aids) including undergarments, pullovers, and front-opening shirts and blouses and managing zippers, buttons, and snaps.</p> <p>0 = Able to get clothes out of closets and drawers, put them on, and remove them from the upper body without assistance</p> <p>Measure does not separate out use of assistive devices.</p> <p>1 = Able to dress upper body without assistance if clothing is laid out or handed to the patient</p> <p>2 = Someone must help the patient put on upper body clothing</p> <p>3 = Patient participates but requires other person</p> <p>4 = Patient depends entirely upon another person to dress the upper body</p>

NOTE: IRF-PAI = Inpatient Rehabilitation Facility Patient Assessment Instrument; MDS = Minimum Data Set; OASIS = Outcome and Assessment Information Set.

An important issue in this demonstration is the potential variation in payment levels and outcomes across settings for the same type of patient. To examine variation in payment levels and outcomes, it is necessary to be able to assess patient severity and needs in a consistent manner between settings that allows for adequate levels of risk adjustment. These are particularly important issues for the more expensive cases, such as patients who need ventilator weaning. The acuity within this patient group may vary, and it is unclear whether each setting provides equivalent resources or has similar success rates in weaning these patients from the ventilator. The same issue applies to other types of medically complex cases.

Understanding patient status across settings is also important for a variety of patient types. Rehabilitation cases (e.g., total knee and hip replacements) are treated in multiple settings where the average payment levels per stay vary. For example, prior RTI analysis suggests that certain conditions, such as DRG 249 (Aftercare, Musculoskeletal System, and Connective Tissue), DRG 012 (Degenerative Nervous Disorders), or DRG 462 (Rehabilitation) may be admitted to both LTCHs and IRFs in nearly equal numbers (Gage et al., 2006). However, the average payment per stay and per day in the LTCH for DRG 249 is greater than the IRF case, even if the IRF admission receives an additional outlier payment. The choice of treatment setting may be due to differences in patient acuity, but this is currently unclear. Standardized patient severity and resource use measures, such as those collected in this project, will allow the evaluation of the extent to which superficially similar populations admitted to multiple types of settings differ in their medical, functional, and cognitive acuity; the extent to which treatment intensity and outcomes differ for similar patients; and, finally, the extent to which different providers are being paid different rates for the same bundle of services.

The need to compare across settings becomes even more important when provider supply variations are considered. Although HHAs and SNFs are widely available, the presence of an IRF or LTCH varies geographically (Gage et al., 2008). Equitable case-mix adjustment values for equivalent levels of care will help ensure that access to cost-effective services is available regardless of variations in provider supply. Better information is needed on patient acuity and the resources used during a stay to allow standardized outcomes analysis and to determine whether these are comparable cases and whether the results of the care applied to comparable cases are equivalent among settings.

The ability to compare risk-adjusted outcomes is a critical need that can be addressed by the implementation of standardized patient assessment items across care settings. Similar types of care and services can be delivered at different PAC provider settings, depending on patient characteristics and the availability of services. However, comparisons of outcomes and quality across these settings are difficult because of the lack of similar measures in each setting to compare the actual outcomes.

2.3 Inconsistencies in Case-Mix Systems and Unintended Incentives

Payments across PAC settings differ considerably, even though the clinical characteristics of the patients and the services delivered may be very similar. The differences in payment among settings can lead to inefficient patterns of care within an episode. In response, CMS has been investigating ways of moving toward consistent and integrated payment approaches. Understanding and comparing the populations served in each of the four settings and the care they receive through the data collected in this demonstration will be an important step toward developing more rational payments and moving away from a “silo-based” approach.

Currently, Medicare uses different PPSs for each of the four PAC providers, each with its own case-mix groups, payment units, associated payment rates, and incentive structures, in addition to variations in eligibility criteria, statutory restraints, and conditions of participation or payment. Each of these systems measures case-mix complexity, but each uses a unique set of items to measure the concepts, making it difficult to compare severity, costs, and outcomes across settings. Despite having different case-mix measurement systems, these four types of

settings do not treat entirely unique populations; many types of PAC conditions are treated in more than one setting. The current Medicare payment methods for PAC providers are designed as independent systems that measure within-setting variation but are limited in the extent to which they handle the potential overlap in case mix or the complementary nature of the services across an episode of care. More importantly, the variability in case-mix measurement and payment methodologies, including both units and adjustment approaches, makes it difficult to compare patient or facility cost differences in a standard way across settings. The payment systems used in the four PAC systems are complex and have many nuances and restrictions. The following represents a simplistic overview and should not be considered a complete description. Briefly, the four systems differ in the following ways:

- SNF patients, admitted after following a qualifying hospital stay, receive a case-mix-adjusted per diem payment that is reset periodically during the stay in the SNF. Payment is based on case-mix or resource utilization groups (RUGs), which are derived from data collected from the SNF patient assessment tool, the MDS. Each RUG has two weights—one each for nursing and ancillary (primarily therapy) costs. The weights are applied to the national per diem rate to create case-mix-adjusted payments per day.
- HHA patients are not required to have had a prior inpatient stay; in fact, a substantial number of HHA patients are direct admits from the community. HHAs are paid on an episode basis that covers a 60-day period. Episode payments are case mix adjusted using the home health resource groups (HHRGs). Payments are adjusted for medical conditions, certain resource utilization patterns, ADL impairments, and the episode's timing in the sequence of patient episodes. Information used to calculate the HHRGs is obtained through the OASIS, the required patient assessment tool used by HHAs.
- IRFs are paid under a discharge-based payment system that adjusts for individual case-mix complexity. IRF payments are based on case-mix groups, which reflect etiologic conditions, functional and cognitive impairments, age, and comorbidities. Case-mix groups are derived from the IRF standardized assessment tool, the IRF-PAI.
- Finally, LTCHs are paid on a case-mix-adjusted discharge basis. LTCHs use the same Medicare Severity Diagnosis Related Group (MS-DRG) system as inpatient prospective payment system (IPPS) hospitals, but the weights are adjusted to reflect the variation in case mix within LTCHs.

In addressing the issue of unintended consequences of independent, silo-based payment systems, CMS will need to examine a variety of issues and determine whether some differences among settings are necessary and warranted. This study represents a starting point for this examination. Important considerations in moving toward more rational payment systems include, but are not limited to, those discussed below.

2.3.1 Consistent Measurement of Items Included in Case-Mix Systems

The payment systems for IRFs, SNFs, and HHAs contain many similar types of measures even though these settings may not measure items in a consistent manner. One of the most important contributions of this project will be the information it provides related to consistent measurement of patient severity across the different PAC settings. For example, the current IRF, SNF, and HHA payment systems all include measures of functional impairment. Standardizing the way function is measured through the use of the Continuity Assessment Record and Evaluation (CARE) tool will be a positive step toward consistency and transparency in the payment systems.

2.3.2 Nonrepresented or Underrepresented Patient Severity Measures

The differences in case-mix systems used in the PPSs are based on historical differences in system development rather than on research examining case-mix differences in each of the settings. The current payment systems may not include all types of patient-level severity adjusters that are associated with greater resource needs. For example, the MS-DRG system used in LTCHs is based on claims information and does not have as an option the inclusion of patient acuity measures from a standardized patient assessment tool. As a consequence, the LTCH PPS uses diagnosis-derived measures of medical complexity and surgical procedures but fails to account for functional or cognitive complexity. This study provides an opportunity to examine the impact of these factors through the use of patient assessment information collected on the CARE tool.

2.3.3 Types of Patient Costs Modeled

The current PAC payment systems differ in whether they attempt to predict patient-specific costs as a whole or whether they break costs into component parts. The approach used in the PAC-PRD analysis separates routine costs into three groups: (1) routine/nursing resource costs, (2) therapy resource costs, and (3) nontherapy ancillary costs per patient. By modeling these components separately, the analysis provides insight into the extent to which each component is important in the different settings, whether they are associated with similar or different patient factors, and whether these factors vary in predicting costs associated with each component.

2.3.4 Unit of Payment

The choice of the payment unit is a critical decision in the development of a payment system. There are three basic choices: day, stay, and episode. **Table 2-2** presents some of the advantages and disadvantages of each unit of payment approach.

**Table 2-2
Selected advantages and disadvantages of payment unit alternatives**

Unit of payment	Advantages	Disadvantages	Methods of avoiding disadvantages
Per diem: institutional PAC day/ home health visit	<ul style="list-style-type: none"> • Provides maximal insurance to providers against unexplained high LOS in a PPS. • Reduces incentives to discharge patients prematurely. 	<ul style="list-style-type: none"> • Reduces incentives to discharge patients at the earliest appropriate stage of their stay, which can increase program costs and potentially reduce the quality by increasing the risk of infection due to longer hospital stay. 	<ul style="list-style-type: none"> • Declining block pricing based on observed marginal costs can reduce incentive to hold on to patients longer than necessary.
Per discharge: institutional stay/ home health episode	<ul style="list-style-type: none"> • Reduces payment amount of risk to Medicare. • Encourages providers to discharge as early as medically reasonable. • Consistent with IPPS, IRF, LTCH, and HHA PPSs. 	<ul style="list-style-type: none"> • Providers may respond to incentive by discharging patient prematurely, resulting in lower quality care to patients, increasing rehospitalization rates, and increasing the use of subsequent PAC, resulting in greater program costs. • Puts providers at risk for LOS differences not explained by case-mix adjustors. 	<ul style="list-style-type: none"> • Cost outlier payments can reduce provider risk for high-LOS/high-cost patients. • Short stay and transfer adjustments reduce program risk for “early discharges.”

(continued)

Table 2-2 (continued)
Selected advantages and disadvantages of payment unit alternatives

Unit of payment	Advantages	Disadvantages	Methods of avoiding disadvantages
PAC episode	<ul style="list-style-type: none"> • Gives clinicians the incentive to determine most cost-effective appropriate mix of post-hospital services. • Limits the program’s costs for potentially substitutable service sites. • Removes program payment incentives for site-of-care choices. • Creates an incentive to discharge to lower cost, downstream providers earlier with expenses handled within the bundle rather than incurring additional program expenses. 	<ul style="list-style-type: none"> • Very different method of payment than current systems, which will be complex to undertake. • If episode-payment sharing rules are not regulated, providers must negotiate payments to each other. • Interim payments may be necessary if episode payment affected by “downstream” diagnoses. 	<ul style="list-style-type: none"> • Administratively determined episode-payment splitting rules can be implemented as transfer-in and transfer-out payments in existing PPSs but may be complex to specify. • Episode payments could be based on average costs of total resources needed to achieve expected post-hospital discharge status (medical and functional), with adjustments for high-cost or short-stay outliers.

NOTE: HHA = home health agency; IPPS = inpatient prospective payment system; IRF = inpatient rehabilitation facility; LOS = length of stay; LTCH = long-term care hospital; PAC = post-acute care; PPS = prospective payment system.

Per-discharge payment, also known as per-stay payment, limits the financial risk to CMS and is useful in settings with discretion in the length of stay (LOS) and the termination of care. General acute hospitals, LTCHs, and IRFs are paid on a discharge-based method, which limits the Medicare program’s liability for each stay but also provides an incentive to inappropriately shorten the LOS and discharge the patient “early” to the next, less-intensive level of care. Policies such as short-term outlier policies are used to mitigate the impact of this incentive.

A stay-level approach is not possible in home health because of its nature as a noninstitutional setting. The HHA PPS uses a hybrid approach, a 60-day episode that can be

effectively extended by initiating a new episode. This gives the HHA a bundled payment for a 60-day period in which it has some discretion over the appropriate number and mix of services. The HHA episode provides some level of protection to CMS from the discretionary nature of visits, especially later in the 60-day period.

A per diem payment unit provides greater risk to CMS by allowing payments to increase with LOS. A per diem approach is most useful with services that are not considered to be discretionary and when CMS is concerned about the impact for incentivizing early discharge. SNF payments are based on a per diem scale and are designed to be adjusted as the patient's health improves. Multiple assessment periods allow the case-mix group used for payment to be reassigned during the course of the treatment. The per diem methodology also reduces the incentive for the SNF to discharge the patient early to avoid additional costs.

Another alternative is to bundle the PAC services into a PAC payment episode. This allows the costs that are most likely to be substitutable to be aggregated as one payment unit and reduces the program risk for different episode compositions. The idea of a bundle is attractive because of the potential for cost sharing and the possibility of consistency in paying for resources needed to provide care. The issue of how to practically implement this approach is complex. To understand some of the ways that patient acuity factors can be used as predictors of episode costs in addition to their use as predictors of within site of care resource use, CMS partnered with the Office of the Assistant Secretary for Planning and Evaluation (ASPE) in an ASPE-sponsored project designed to use the CARE data collected in the PAC-PRD project to predict PAC episode payments. In addition, the CMS Bundled Payments for Care Improvement Initiative invites providers to propose different bundling approaches within certain broad parameters that vary in the extent to which they cover PAC services and the nature of the interaction and involvement among PAC providers.

The PAC-PRD analysis offers the opportunity to revisit the decision of the appropriate unit of payment in the different PAC settings. One consideration related to the unit of payment is whether to implement a consistent unit of payment across all PAC payment systems or instead to maintain the current approach of using different payment units across PAC PPSs. Retaining the existing units of payment would be least disruptive but would complicate the task of coordinating payment across systems. Advantages of a consistent approach include that it is (1) easier to ensure consistency across providers in incentives to discharge, (2) easier to ensure consistency in paying for resources needed to provide care, and (3) easier to translate estimated payment models into payment systems. Disadvantages of a standardized unit of payment across a continuum of care may be that (1) using the same type of payment unit may not be consistent with desired incentives for using different types of services, and CMS may wish to use broader bundling units for those types of services where the expected service units are predictable and cannot be substituted in an alternative setting; (2) even with common case-mix measurement, within-payment-unit variation in cost may differ across providers, resulting in differences in ability to "cherry-pick"; and (3) the use of a standardized unit of payment may not reflect existing practice patterns in all settings.

Within the PAC-PRD analysis, models are assessed at both a stay level and a day level in an effort to provide information that can be used in a flexible manner. Because home health

occurs in a noninstitutional setting, a 60-day episode was considered to be equivalent to an inpatient stay.

2.3.5 Service Use Measures

The current PAC payment systems differ in whether they incorporate service use measures in their case-mix systems. The inclusion of service use is problematic because it is thought to be “gameable” or subject to discretionary changes not related to the actual care needs of the patient. It is preferable to include the patient factors that would lead to the need for treatment (such as the functional impairment leading to the need for therapy use) rather than the fact that treatment (e.g., therapy) occurred. In examining predictors of resource use, the PAC-PRD analysis avoided measures of service use within the setting where possible and focused on less discretionary types of services, such as ventilator or hemodialysis use.

2.3.6 Course of Treatment Perspectives

The absence of a standardized case-mix measurement system that could be applied in all medical and functional rehabilitation settings has restricted the Medicare program’s ability to move toward an episode-of-care approach to paying for services. Instead, much of the research has been restricted to silo-specific approaches using the case-mix tools that are designed for each system to examine costs and quality within each site of care. Although each PPS may provide appropriate incentives to discourage/encourage different length stays at different levels of intensity, they fail to account for the impact of different service mixes across a broader episode of care. About 34 percent of all acute discharges are discharged to PAC, and among them, 24 percent will use two PAC sites during an episode of care, whereas another 4.6 percent will use three or more settings during an episode (Gage et al., 2008). A substantial number of these patients will be readmitted to the hospital. The likelihood of using multiple providers during an episode is common across many diagnoses and increases with the severity of illness or case complexity factors.

The current payment systems could be improved in their ability to address potential cost shifting between providers. Currently available approaches to addressing cost shifting include such policies as the transfer adjustment policy in the IPPSs and the short-stay transfer policies in IRFs and LTCHs. PPSs may also include adjustment for prior service use such as in the HHA PPS, which includes an indicator of the HHA episode number. Indicators of prior service use may serve as a proxy for patient acuity factors that vary by prior use, but it is preferable to use direct measures of acuity if possible.

The issue of hospital readmission has received an increasing amount of attention. Both LTCHs and IRFs are financially responsible for an inpatient stay if their patient has an interrupted stay or returns to the hospital for a limited number of days before returning to them. Home health providers are also subject to regulations regarding how to handle hospital stays that occur within their 60-day episode.

2.3.7 Incorporating Issues of Value Into Payment

Although CMS has an established history of assessing quality in SNF and HHA settings, it is just beginning the process of standardizing quality assessment in LTCHs and IRFs in

response to Section 3004 of the Affordable Care Act. Reporting and understanding quality is an important component of the process of incorporating issues of value and quality into payment mechanisms. The principles of value-based purchasing have not yet been incorporated into the various PAC payment systems.

The collection of consistent patient information at admission and discharge from PAC settings will allow CMS to measure patient factors and outcomes across settings. The additional collection of information at discharge from the hospital will allow outcomes and severity to be examined on a trajectory-of-care basis instead of within a site of care. The systematic implementation across provider settings of standardized data from the CARE item set will lead to better analysis and comparison of beneficiaries' clinical complexity, severity, and outcomes and will allow for the further development of value-based purchasing initiatives within PAC settings and potentially across trajectories of care.

2.4 Other Rules and Factors

Numerous other factors make up a payment system, including the structure of low- and high-cost outliers, what services are included under consolidated billing (including which specific types of ancillary services are covered), the specific nature of patient cost sharing, and Medicare coverage and payment requirements. Systems that incorporate assessment data also have the additional complication of using different assessment windows. If a unified payment approach goes into effect, standardized measurement times will also need to be considered. These factors are not directly addressed in this report but are raised in the context of moving toward paying for and properly incentivizing PAC services.

SECTION 3

DEVELOPING STANDARDIZED MEASUREMENT APPROACHES: THE CONTINUITY ASSESSMENT RECORD AND EVALUATION (CARE)

This section reports on the development of a standardized set of assessment items for measuring medical, functional, cognitive, and social support factors in the acute hospital, long-term care hospital (LTCH), inpatient rehabilitation facility (IRF), skilled nursing facility (SNF), and home health agency (HHA) settings as directed by Section 5008 of the Deficit Reduction Act of 2005. These data were used in the Post-Acute Care Payment Reform Demonstration (PAC-PRD) to collect case-mix data in each of the five settings. This section provides a brief overview of the work undertaken in this areas. For additional information, please refer to *The Analysis of the Reliability of the Items in the Continuity Assessment Record and Evaluation (CARE) Item Set Report* (Gage et al., 2010).

Each of these settings currently has some set of assessment items used at intake and throughout a patient's stay to document health status. Most assessment tools measure the same underlying concepts of patient acuity, but they may use different items to measure these concepts. The Medicare program currently mandates three different assessment tools to collect health and functional status information on patients in IRFs, SNFs, and HHAs. The required tools include the Minimum Data Set (MDS) in SNFs, the Outcome and Assessment Information Set (OASIS) in HHAs, and the Inpatient Rehabilitation Facility-Patient Assessment Instrument (IRF-PAI) in IRFs. The data are used to adjust payments and quality measures to reflect the complexity of the individual patient being treated. General acute hospitals and LTCHs each collect data on similar concepts at intake and during patient stays, although they may not use a standardized set of assessment items across providers and, for certain items, may have the information in medical notes rather than standardized assessment items. The exact assessment tools used typically differ by hospital or corporation, making it difficult to share concise information about patients as they transfer between settings.

In addition to the specific questions being asked, the current assessment processes differ in other ways as well, even among the three federally mandated assessments. Each tool uses different assessment windows resulting in the patients being assessed at different times during their treatment. Even the defined period for an admission assessment differs across the three tools, making it difficult to compare severity, outcomes, and cost across providers over time. As well as variable definition differences, the MDS, OASIS, and IRF-PAI have incompatible data formats; thus, it is difficult to share data electronically across levels of care. Within settings that have integrated data systems across different levels of care, the three federally mandated tools are either excluded or have to be incorporated by the software vendors into the existing system.

3.1 Introduction to the CARE Tool

This section discusses the development of the CARE tool, which contains a set of interoperable data items that can be exchanged using Health Level Seven International (HL7) standards and that were developed with the input and consultation of the clinical communities serving Medicare beneficiaries in the general acute, LTCH, IRF, SNF, and HHA settings. The items were tested for reliability in each of the five populations and can be used to replace similar, nonuniform items on existing assessment tools.

3.1.1 Stakeholder Input

The development of CARE was a multipronged effort that involved extensive input from numerous stakeholders, experts, clinical groups, and information technology thought leaders. RTI worked closely with the Office of Clinical Standards and Quality and the Office of Information Systems at CMS and their colleagues in the Innovation Center and in the Center for Medicare to address quality, payment, research, and survey and certification needs. Key stakeholders from the five different research and clinical communities associated with acute and post-acute care (PAC) services identified the core set of items that are needed to measure patient complexity, regardless of site of care. Input was collected through numerous stakeholder meetings, including several open door forums (ODFs) and technical expert panels (TEPs), as well as smaller, ongoing discussions with members of the different national provider associations.

3.1.2 Item Selection

CARE tool items were limited to those needed for payment or quality monitoring. The CARE effort attempted to use standardized versions of the currently mandated assessment items in the Medicare payment systems, including those in the IRF-PAI, MDS, and OASIS instruments. Items from the existing MDS and OASIS tools that were used only for care planning were excluded from CARE. Items identified by the acute hospitals and LTCHs for assessing patient severity at admission or during a stay also were included in the CARE item set. Most of the items in the CARE item set are currently typically recorded in patient charts, although the format, formality, location of the data in the record, and designated individual(s) or clinician(s) on staff who collect the data (e.g., nurse, therapist, case manager) may vary.

3.1.3 Response Rate

CARE data were collected in the PAC-PRD between 2008 and 2010, the 3 years of the demonstration. Over 53,000 assessments were collected in nearly 200 settings, including acute hospitals, LTCHs, IRFs, SNFs, and HHAs.³ An additional 455 assessments were collected to test inter-rater item reliability of the standardized CARE items, and an additional 550 assessments were collected in a second reliability approach to test reliability across disciplines and settings. A complete report on the development of the CARE items is available (Gage et al., 2008). A second report, *Analysis of the Reliability of the Items in the CARE Item Set* (Gage et al., 2010), presents the results of the reliability tests of these items when applied in each setting although a summary of each is presented here.

This section summarizes the development of the standardized assessment items, including the item selection process and the reliability testing for the items collected in the PAC-PRD. Because the demonstration involved clinicians who practiced at many different levels of

³ The data collection was divided into two major phases: an initial phase covered under the DRA of 2005 and a supplemental phase covered under MMSEA (Medicare, Medicaid and SCHIP Extension Act) of 2007. The two phases had different analytic focuses and thus differences in the markets and providers chosen for inclusion. Depending on the analysis, the data used in this final report may include only that which was collected in the initial data collection effort or information from both initial and supplemental collections. See each chapter for more information on the specific samples used in the analysis.

care or treated patients at different levels of complexity, the CARE tool included multiple versions of some concepts. This allowed empirical testing to determine which measures of a concept had the best reliability across the spectrum of settings. Although each item was selected from items that were already validated in one setting, few had been tested in more than one setting. The reliability tests had to examine whether a specific item may be limited in its ability to capture the complete range of severity when applied to a different population or in a different level of care. This issue was particularly a concern with capturing the full range of functional performance from the very impaired to the very fit, although it also applies to medical status items, such as pressure ulcers.

Development of the CARE tool was successful in terms of the following:

- CMS achieved its goal of developing a standardized assessment instrument that is useful; clinically relevant; grounded in scientific evidence; flexible for easy, rapid accommodation of future clinical and technological advances; electronically based on federally and nationally recognized standards for interoperability across settings; and generally supported and accepted by stakeholders.
- CARE lays the groundwork for enabling providers to use a uniform set of data elements to assess beneficiaries' progress and outcomes achieved in relation to resources used in various health care provider settings. CARE successfully meets the legislative directive to collect data predictive of outcomes and resource utilization that can guide quality and payment policy development. Additionally CARE provides a standardized data collection vehicle for measuring beneficiaries' health and functional status longitudinally across settings and episodes of care. This will enhance clinical communication by standardizing the language used to measure patient severity and allow electronic exchanges that can facilitate better care coordination.
- CARE successfully moves CMS and providers forward from the use of multiple incompatible assessment instruments to one standardized set of clinically relevant data that applies federally and nationally recognized health information technology (HIT) standards. Use of broadly adopted HIT standards will allow for the safe, secure, electronic exchange of critical health information among authorized users.

3.2 Guiding Principles of CARE Tool Development

The CARE tool's development was based on certain guiding principles. As laid out in the initiating legislation, the CARE tool needed certain characteristics:

- The CARE tool should be designed to collect standardized information at discharge from acute hospitals and at admission and discharge from the four PAC providers: LTCHs, IRFs, SNFs, and HHAs.
- The CARE tool items should inform payment policy discussions by including measures of the needs and the clinical characteristics of the patient that are predictive of resource intensity needs.

- The CARE tool items should inform the evaluation of treatment outcomes by including patient-specific factors that measure outcomes and the appropriate risk adjustment thereof. Outcomes should include but not be limited to measures of functional status.
- The CARE tool items should document clinical factors associated with patient discharge placement decisions for the purposes of allowing the clinicians treating the patients to make appropriate discharge placement decisions.
- The CARE tool should be appropriate for collecting standardized patient assessment information as a patient is transferred from one setting to another and, by standardizing how information is collected, foster high-quality, seamless care transitions.

Item selection was based on several overriding principles:

- Sensitivity to data collection burden. Selected concepts and items were restricted to those that were typically already in use for payment or quality monitoring purposes or would improve these efforts.
- Consideration of the reliability and validity of items. Items included in the Federal set needed to be reliable and valid measures of the concepts they were intended to measure.
- Breadth of application to minimize floor and ceiling effects. Certain items in the existing tools were limited by floor and ceiling effects in their ability to explain variation across patients having a broad range of severity within the measured clinical characteristics as found in the PAC populations, but these items should reduce those effects.
- Minimization of “gameability” or incentives that might encourage provider behavior that is inconsistent with best practices for patient outcomes and care quality.

Overall, the development work had to build on the current scientific knowledge, incorporate the guidance provided by the five different measurement and clinical communities, and minimize provider burden in collecting the data. Item development was based on extensive participation of stakeholders throughout the process. CMS invited provider associations from each of the five levels of care to nominate participants for different TEPs. The first panel asked the clinical community to define the most important concepts to include to measure differences in patient severity or factors that would affect resource needs and outcomes in their populations. The second TEP included measurement experts from each of the five provider communities to discuss the potential items that could be used to measure the proposed concepts. Extensive literature exists on the reliability and use of items in their respective areas and their effectiveness within a level of care, but this TEP was asked to consider issues related to developing uniform measures across settings and identifying the best approaches. A pilot test was held to test the proposed items and the data collection process in each of the five levels of care. The resulting data were presented to a third TEP to further refine the proposed item set.

In addition, RTI and CMS held small group meetings with a variety of association members throughout the process to review materials and receive feedback on the tool. The feedback was incorporated into early tool refinements and the design of the data collection process. RTI and CMS sought feedback particularly on the relative ease of completing each item within each provider population and also on practical considerations including the training sessions and the Web-based data entry/submission system. Clinical input also led to HIT refinements that helped with the design of screen content and methods for moving between sections of the tool.

The design also needed to take into account operational feasibility. While IRFs, SNFs, and HHAs already had procedures in place to submit their assessment tools to CMS, general acute hospitals and LTCHs had not been submitting assessment data to CMS. Consideration of their current assessment practices was used to determine the feasibility and best approaches for collecting the CARE items electronically. Operational procedures were discussed during the demonstration site setup calls so that data collection practices at the individual provider sites were as consistent with actual assessment practices as possible. The one caveat was that the data items needed to be consistently assessed across all participants.

The CARE item set also needed to recognize provider burden. Two types of items were included—a core set to measure severity (or presence of a factor) on any beneficiary receiving treatment and a supplemental set that provides standardized items to measure the severity of conditions when present. The core items provided a select set of data on patient medical complexity, functional impairment, and discharge status. The supplemental items provided standard language for measuring a set of items that refined the severity of conditions present. For example, all patients were assessed on the one screening item for pressure ulcer, but the rest of the pressure ulcer items measuring numbers and severity were only completed for those who had a stage 2 pressure ulcer or worse. Using a core/supplemental item approach allowed standardization of the language clinicians use across sites of care, while minimizing the number of items assessed on individual patients. Only the most complex patients were assessed on the total item set; the healthiest populations' assessments were limited to core items.

This first generation of CARE items targets basic core and supplemental items for measuring frequently occurring conditions in the Medicare populations, such as medical, surgical, and functional conditions. In the future, standardized subsets of CARE data, or modules that are more specific to a particular condition or provider setting, could be drawn from the registry storing the standardized CARE library of elements and concepts. This approach will allow item modules to be added in the future as more of the clinical items used in quality monitoring and survey and certification become integrated or, alternatively, allow items to be merged with other data sets. For example, the CARE data set could be merged to the MDS or OASIS files to incorporate care planning items associated with individual patients but not relevant for payment or quality purposes. Additionally, standards-based items could be added to capture individual patient preferences for care treatments, along with items that measure the degree to which individuals' preferences and goals have been met. In essence, CARE has been designed to evolve over time to incorporate a broader range of items that address patient-centered care planning, quality measurement and reporting, and other emerging needs.

Last, the CARE items were designed to be an interoperable item set that can change as medicine changes. The CARE vehicle contains HL7-based electronic components that will allow the exchange of data across different systems. CARE provides a dynamic framework for housing a standard set of items that can be used across the Medicare program, stored in an item library, and exchanged through interoperable data exchanges. Each item meets the national standards for health data exchanges as set by the Office of the National Coordinator. This framework will allow *standard* items to be used without requiring that all providers collect every item; instead, the individual items can be specific to one setting or another, as required by the program or needed by the provider. By providing interoperable, standardized items, a national standard is in place that will ease electronic transfers of data across providers and among authorized parties, such as the Medicare program.

3.3 Development of the CARE Tool and CARE Tool Items

The Deficit Reduction Act of 2005 mandated that the PAC-PRD be in place by January 2008. This timeline required that the CARE tool be ready within a 14-month window. Given that the charge was to build on the current science, develop a consensus regarding the most appropriate measures from each field, and test the tool in each of the five settings, this work progressed on a steady schedule.

Recommendations for items to include in the CARE tool were based on a critical review of the current assessment tools used in each setting, incorporation of proposed changes in the MDS 3.0, the OASIS-C, and the IRF-PAI QI and consideration of the World Health Organization's development of the International Classification of Function (ICF) model and other measurement efforts in the fields of critically complex medicine, wound care nursing, and related areas. The Institute of Medicine's (IOM's) six key aims to provide safe, effective, efficient, patient-centered, timely, equitable patient care were central to CARE's development. Additionally, to be considered, items had to have been validated with at least one population and be free of copyright restrictions. RTI brought together a wide range of clinicians, providers, and researchers to identify the necessary concepts, review existing measures in each field, and develop a consensus regarding the best measures of each concept. Items were selected based on their importance for measuring patient severity, resource needs, or outcomes and their ability to detect differences across the range of PAC patients. Input on the selection of the core items appropriate for measuring baseline complexity (medical, functional, and cognitive complexity) and on the best measures of those concepts was provided by teams of clinicians representing each of the five levels of care, including acute hospitals, LTCHs, IRFs, SNFs, and HHAs.

3.3.1 Defining the Domains

The first step in developing the CARE tool was to examine the domains common to each existing assessment tool and determine which types of concepts should be included in this standardized item set. The tool needed to effectively measure patient severity factors that would predict the need for different types of treatments or resources or measure outcomes. Based on the 2006 report *Uniform Patient Assessment for Post Acute Care* (Kramer and Holthaus, 2006), five primary domains were selected. The first four domains—medical, functional, cognitive, and social support—are common to most medical assessment tools regardless of site of care. The fifth domain—transition items—was identified as important for improving quality of care. By improving information transfer between sites, avoidable hospitalizations and other adverse

conditions can be prevented. Providers from all PAC levels of care were involved in identifying the necessary items.

The first four sets of domains were identified as key to distinguishing different resource needs in each setting and potentially affecting outcomes, if present. Each domain has a small set of core items applicable to all patients and a set of supplemental items. The majority of items is supplemental and is used to measure severity of a condition only if a condition is present. Hence, not all factors are assessed on all patients, but those that are relevant are collected in a standard way. The four domains include the following:

- **Medical Status/Clinical Complexity.** These items measure patient medical status and include factors defining complexity in terms of medical diagnoses, resource use such as procedures or major treatments received during stay (e.g., ventilator weaning, hemodialysis), medications, skin integrity (number and size of pressure ulcers and locations and presence of other wounds), and physiologic factors (e.g., vital signs, laboratory results, blood gases, pulmonary function).
- **Functional Status.** These items include screening items on impairments (e.g., bladder, bowel, swallowing, vision, hearing, weight-bearing, grip strength, respiratory status, and endurance) as well as measures of self-care, mobility, and safety-related functions (medication management, phone management) and other items relevant to less impaired populations.
- **Cognitive Status.** These items target memory/recall ability, delirium/confusion (some of which may be short term related to current medications or longer term, which may complicate rehabilitation therapy), behavioral symptoms including those that are self-injurious (pulling IV lines) or directed toward others, signs of depression or sadness, and presence of pain, which may affect patients' engagement and outcomes.
- **Social Support Factors.** These items target social support issues, including information on structural barriers, living situations, caregiver availability, and the need for assistance, as well as issues related to discharge complications.

Together, these four domains provide a comprehensive overview of a patient. For healthier patients, fewer items are relevant. For more complex patients, the CARE items offer standardized versions of information already collected on those types of patients. The fifth domain, transition items, included items that are important for the transfer of information between facilities but were not otherwise captured, such as information on allergies.

3.3.2 Forming Clinical Workgroups

The initial RTI work was done by a large team of clinical staff from various backgrounds, including geriatric medicine, pulmonology, infectious disease, internal medicine, physiatry, medical and rehabilitation nursing, occupational therapy, physical therapy, epidemiology, intensive care, and public policy. Team members included staff from RTI as well as subcontractors from the Rehabilitation Institute of Chicago, Evanston Northwestern

Hospital/National Institutes of Health (NIH) Patient Reported Outcomes Measurement Information System (PROMIS) team, Northwestern University, and consultants from the University of Pennsylvania, Case Western University, RAND/VA, and the Visiting Nurse Service of New York. Extensive input was also provided by our pilot test sites, including RML Specialty Hospital, Edwards Hospital, Rush Copley Hospital, Marianjoy Rehabilitation Hospital, ManorCare Corporation, and the Visiting Nurse Association of Fox Valley. Clinicians represented each of the five levels of care: acute, LTCH, IRF, SNF, and HHA.

Four clinical workgroups were established, each responsible for a different conceptual domain. (Care transitions were handled within the medical acuity group.) Representatives from all five levels of care participated in each workgroup. The clinical teams focused on item selection and the goal of each recommended item in preparing materials for TEP review. Response burden was a constant criteria applied in each workgroup. The final list of items proposed to the TEPs was restricted to those measuring patient treatment needs or outcomes. Each item had to be justified for its inclusion in the CARE tool. (See Gage et al. [2008] for further information).

3.3.3 Selecting Items for Use in the CARE Tool

The four workgroups were asked to identify the best items under each domain that could be applied across the range of health and impairment levels treated in these settings. Although each of the current assessment tools measured similar concepts or subsets of concepts in each setting, they used different items to measure the concepts. The CARE items are the result of these discussions and represent standardized versions of the identified item. The workgroups received input and oversight throughout this process from the TEPs, provider and stakeholder input, and CMS review.

Many of the items that were considered for inclusion are the same as those in the MDS 3.0 and OASIS-C because these two instruments were going through re-evaluation at the same time and this work was done in collaboration with that effort. At the same time, the CARE tool has many fewer items than the MDS or OASIS because the two setting-specific tools also have care planning items that are not necessary for cross-setting measurement of severity.

The CARE tool also built on the IRF-PAI tool in identifying important concepts or domains for measuring severity in the populations needing physical rehabilitation services. Input from the field was used to refine measurement approaches that identified an impairment or level of independence but improved measurement of function across populations. Similar inputs and revisions were based on recommendations from experts in the pressure ulcer measurement community, including the National Pressure Ulcer Advisory Panel and others. The CARE tool also has a few items that measure severity in the more medically complex populations treated in inpatient settings, such as acute hospitals, LTCHs, and IRFs. These items are based on those currently used in the acute and LTCH intake or assessment processes. Last, certain factors were important for understanding discharge options and safety. These were largely based on the input of the home health and case management fields. The result is a standardized set of items measuring medical, functional, and cognitive deficits and standardizing discharge-related items. The versions of the CARE tools used in data collection are available on request.

3.3.4 Basic Organization of the CARE Tools

The result of the four clinical workgroups led to development of a CARE tool that was used in two rounds of pilot tests. The results from the pilot test were used in TEPs and resulted in revised versions of the CARE tool that were subsequently published in the *Federal Register* for public comment.

The CARE items provide standardized approaches for measuring medical, functional, and cognitive status across settings and over time. In effect, it provides a virtual electronic health record for a Medicare beneficiary. As in any medical record, some items will not be relevant and will not therefore be completed beyond a screener question. But the system standardizes the items that are used across the five settings to define the patients' medical, functional, and cognitive complexity.

In addition to the standardized items to measure each concept, the CARE tool also standardizes the assessment periods to define the window of time that reflects a patient's admission period or discharge period. Consistent assessment windows (e.g., "x days before or following hospital discharge") were needed to allow comparison of patient acuity at the same point in time, regardless of subsequent service sites. Currently, each mandated measurement system uses different assessment windows to describe patient severity. The IRF-PAI includes data collected during the first and last 3 days of a stay, the MDS collects admission data within the first 5 days of an admission and at subsequent follow-up times, and OASIS data are collected during the first visit, which may vary by when the HHA was able to initiate care, rather than reflect the patient at a specific time period following discharge from the hospital. As a result, each system may be assessing patients at different points in their episode, which will affect the severity ratings found in each tool. The CARE tool established standard assessment observation windows (time frames) across all five settings for time-sensitive data.⁴ The time frames used in CARE were 2-day assessment windows at admission and discharge. These observation windows could be extended by 1 day if the admission or discharge occurred after noon. For the home health setting, assessments were completed during the first and last visits. These observation windows were chosen to allow comparisons of clinical complexity, severity of illness, and functional status at specific points in time across provider settings. Sufficient time frames were factored into the assessment windows to allow adequate time to assess the patient.

The information collected was standardized within and between settings. Where appropriate, measures were also collected consistently between the admission and discharge forms to measure changes in clinical acuity or functional performance. At the same time, some items are only relevant at admission; others are important at discharge, especially if a patient is returning to the community.

CARE tool items were selected with the goal of capturing patient acuity for the entire range of severity: from the comatose patient to the patient about to be discharged from home health without any remaining concerns. As mentioned above, CARE was designed with a small,

⁴ Specific items, such as items measuring acuity or impairments, were identified as time sensitive. Items related to demographics or premorbid status that would not change during the stay could be completed outside of the observation window but before the assessment was finalized.

core subset of standardized items that apply to all patients. CARE incorporates screener questions to allow less clinically complex patients with few issues to be assessed quickly. Greater detail is solicited by collecting additional CARE items on more complex, sicker patients.

One of the major changes made in the transition from MDS 2.0 to MDS 3.0 was the expansion of measures that directly captured the patient's voice through interviews or captured the patient's experience through direct observation of the patient's performance. The CARE tool also sought to capture the patient's voice in the items chosen for inclusion. Both patient self-report and clinical perceptions are included in the tool to the extent possible. The exact manner in which interview items were used in CARE was guided by input from the clinical communities. The clinical communities thought the importance of each differed by domain. For example, the ideal pain measure is based on the patient's perception, whereas mobility or self-care skills were felt to need clinical assessment of the patient's ability.

The CARE item set was designed as a starting point for standardized assessment items across the Medicare program. Additional items or modules can be added in the future, but this work focused on the minimal items needed to measure baseline acuity or quality of care.

3.4 CARE Item Description

The final CARE item set used during this project has nine sections that reflect the domains currently collected in most patient assessment tools or patient intake forms. Some of the items may not currently be included on all intake assessment forms, but most are noted in the patients' charts, at least informally if not uniformly. The items affect how clinicians provide care, including information on premorbid impairment levels and current cognitive complications. This section describes the CARE items in more detail and the use of the items in our analysis.

3.4.1 Administrative Items

The administrative items are basic insurance information items that identify the admitting provider, insurance coverage, and demographic information, such as age. These items are based on current Medicare administrative data collection and related certification procedures. Earlier versions of the item set also included educational levels, but these were omitted for brevity.

3.4.2 Premorbidity Patient Information

The premorbidity items provide baseline data on the patient's preadmission service use in the last 2 months; residential information, including type of residence prior to admission, whether they lived alone, and type of help used in the community setting; structural barriers at home; prior physical and cognitive functional status; use of assistive devices; and falls history. Some version of these items is typically collected at intake in each setting. These standardized versions are based on existing items in the Medicare program. They will be important risk adjusters in measuring outcomes, including the probability of discharge home to the community and expected changes in functional limitations.

3.4.3 Current Medical Information

The current medical information section provides the most important information for explaining medical or level-of-care needs. Patients with greater medical complications need

more intensive settings with higher frequency physician and nursing care. The inpatient settings range in medical intensity from acute intensive care units to acute step-down units or LTCHs to SNFs to HHAs. Each level has declining physician and nurse full-time equivalents. The factors in this section are commonly used in current case-mix systems, such as diagnosis, comorbidities, procedures, and skin conditions or else commonly collected on current assessment tools to determine staffing ratio needs on particular units, such as treatments, and physiologic factors. These items, in combination with the cognitive and functional factors, are important measures of variation in patient acuity.

Many of the items in the current medical information section are taken from the patient's medical record and are organized to be supplemental items that are answered only when the screening item identifies the items' appropriateness for the individual patient. Not all items in this section apply to all patients. Some items, such as primary and secondary conditions, are core measures of illness and are collected on every patient; other items, such as those under the major treatment section, are only applicable to patients having those more intensive treatments. They are predictors of resource use, in terms of nurse staffing and physician frequency needs, potential rehospitalization predictors, and complications in analyzing outcomes. They are also important measures of changes in medical status during an admission.

The last medical section collects physiologic factor information on vital signs, laboratory tests, arterial blood gases, and pulmonary function tests if these tests were conducted; otherwise, they are not applicable to the patient's health status.

3.4.4 Interview Items: Cognitive Status, Mood, and Pain

Stakeholder feedback to CMS underscored the importance of including patient-centered interview items that reflect the voice of the patient. The patient interview items included in this section of the CARE tool are important risk adjusters for analysis of both outcomes and resource needs. Patients with cognitive impairments are less able to communicate with their providers, carry out treatment instructions, and achieve equal outcomes to patients who may be equivalent in terms of medical conditions. These items include an orientation/memory/recall item and a delirium item. These two sets of items were identified by the TEPs as important in all five levels of care but not consistently measured. Delirium was particularly important in the discussions of patients being transferred between settings, and the measure chosen to assess delirium, the Confusion Assessment Method (CAM), had been previously tested in populations at different levels of care. The Brief Interview for Mental Status (BIMS) was chosen as the means to assess cognitive status in CARE. The BIMS measure was being used in the MDS 3.0 and was found to be a strong measure of memory/recall for patients receiving skilled services. An observation-based assessment of cognitive status was used in the event of a patient not being able to be interviewed.

Pain measurement items are also included in this section because, like the cognitive measures, they require patient interviews to document the level of pain and its effect on the patients' treatment. Because these items are interview based, this section includes two sets of items—an interview version—and when a patient cannot be interviewed—an observation-based item measuring the same concept. Patients are asked to report their pain on the standard 0–10 scale used in most hospitals, LTCHs, and IRFs and also asked to report whether the pain limited

their sleep or activities in the past 2 days. This approach allows for better measurement of pain effects across people who may have different pain thresholds. Clinicians complete either the interview or the observational item, although during the demonstration, some clinicians suggested that both items should be completed on every patient.

Two measures of depression are included in this section. Both are interview based but were initially developed by different groups. The first item is the two-item Patient Health Questionnaire (PHQ-2), which asks patients how often over the past 2 weeks, they had low interest or were feeling sad. This item is a modified form of the longer MDS 3.0 item (PHQ-9). The second depression item is taken from the NIH/PROMIS initiative and asks patients to answer how often they felt sad in the past 2 weeks using a 5-level scale with “0” being never sad in the past 2 weeks and “5” being always sad.

3.4.5 Impairments

The impairments section contains a series of screening and supplemental items to identify impairments that restrict a patient’s ability to function but which are not direct measures of functional abilities. These items are important risk adjusters for considering outcomes and resource needs. Included are measures of bladder and bowel incontinence; swallowing abilities; hearing, vision, and communication skills; weight-bearing restrictions; grip strength; respiratory status; mobility and sitting endurance; and use of assistive devices, such as canes, walkers, wheelchairs, and other devices. These types of measures are commonly collected on populations with physical rehabilitation needs, and most are included in the federally mandated IRF-PAI, MDS 2.0, or OASIS tools. Much of this section is screened out for relatively healthy patients with no impairments. But for those who have an impairment, this section provides a standardized item to measure its severity.

3.4.6 Functional Status

The items in the functional status section are performance based and measure the level of assistance needed by these patients at admission and at discharge. Within functioning, we included variables related to the subscales of self-care, mobility, and instrumental activities of daily living (IADLs).

The work builds on the science of the physical rehabilitation field but uses a different approach than the FIM[®] function measures currently in the IRF-PPS. In addition to the FIM[®], the CARE items build on work by Stineman (1996), Jette (1996), and others who have built on measures of the need for assistance that were initiated with the Barthel Index to measure a patient’s ability for self-care or physical mobility.

Although similar to functional performance measures used in the IRF-PAI data collection, the CARE functional items differ in the specific types of performances being examined, the use of a 2-day observation window, the evaluation of the patient’s usual performance (as opposed to the best or worst performance), and the use of a more simplified scoring approach for the scales on which each specific performance is rated. If an activity was attempted, the patient’s performance was noted on the following scale:

6. **Independent**—Patient completes the activity by him/herself with no assistance from a helper.
5. **Setup or clean-up assistance**—Helper SETS UP or CLEANS UP; patient completes activity. Helper assists only prior to or following the activity.
4. **Supervision or touching assistance**—Helper provides VERBAL CUES or TOUCHING/STEADYING assistance as patient completes activity. Assistance may be provided throughout the activity or intermittently.
3. **Partial/moderate assistance**—Helper does LESS THAN HALF the effort. Helper lifts, holds, or supports trunk or limbs but provides less than half the effort.
2. **Substantial/maximal assistance**—Helper does MORE THAN HALF the effort. Helper lifts or holds trunk or limbs and provides more than half the effort.
1. **Dependent**—Helper does ALL of the effort. Patient does none of the effort to complete the task.

The scale used in the functional performance items identifies whether the patient needs assistance to complete more or less than half an activity rather than requiring the clinician to evaluate the need for assistance in terms of quartiles. For those who need less than 50 percent assistance for the activity, the categories are further refined by whether the helper must remain present for safety supervision or cueing or whether they can set up the patient and walk away from them without concerns for safety. For patients who need help with more than half the activity, distinctions are made between the patient who needs total assistance and the one who can do some part of it independently. This coding method was well received by the therapists across all the settings. Anecdotally, the feedback was that more accurate measures can be made with this approach than with determining whether someone needs help with 20 percent, 25 percent, or 30 percent of an activity, all of which result in different group assignments under the current IRF classification system. Therapists also preferred the CARE functional scale to the grosser measures of function in the MDS and OASIS tools, which they felt failed to show patient gains in a meaningful way.

Like the medical section, these function items are divided into core measures of self-care and functional mobility needed to provide baseline information on all patients and supplemental items that will allow more refined measurement of patient ability, given the presence of a limitation in the core items. A wide range of activities were evaluated to address some of the ceiling and floor effects seen in functional performance measures used in the FIM[®], MDS, and OASIS. For the demonstration, providers were instructed to collect functional information on all of the items with the goal of analyzing the patterns of functional performance within and between provider settings and, in the future, potentially reducing the number of items needed to accurately assess functional ability.

3.4.7 Frailty and Life Expectancy

Measures of frailty and life expectancy are also included in the CARE items because they may identify complications that are difficult to assign to a specific medical, functional, or cognitive impairment but can affect one's function level. The concept of frailty is considered

important in measuring geriatric health status (Fried, 1997), but its measurement is difficult and has been poorly defined at this time. The CARE items include a measure based on a mix of factors that may vary by type of patient but together at the individual level suggest the patient is frail.

The CARE measure of frailty is based on the clinician's perception of the patient's overall health status and whether this patient is in a late state of decline. This is an item adapted from the British Gold Standards Framework Programme (National Health Service, 2005). A second item included for its correlation with frailty is the grip strength item included in the impairment section. Impaired grip strength is commonly used as a performance-based measure of the patient's strength in both physician and nursing practices. The absence of strength is commonly thought of as a manifestation of frailty.

3.4.8 Discharge Information

The discharge information section of the tool collects information on patient discharge destination and nonmedical factors that might affect these decisions. Social factors, such as the availability of caregivers and their ability to meet the required level of need, are examined. Patient readiness factors, such as the ability to manage or pay for medications and the availability of transportation assistance, are also documented in identifying the availability and capability of a caregiver following discharge.

Additionally, the discharge section collects information on discharge decision-making issues, including identification of the range of PAC providers considered appropriate by the medical team, the availability of those types of services in the local area, the availability of insurance coverage for these services, and the effects of patient or family refusals for certain types of providers. Understanding the clinician's perceptions of the potential for treatment at alternative settings, and how this varies in different parts of the country and for different types of cases, will provide insight to some of the more complicated factors affecting level-of-care decisions. These items are currently documented in social workers' notes, but they are not consistently recorded nor done so in a standard, comparable way. Feedback from the field, particularly the hospital and SNF communities, suggested they liked the idea of documenting these issues because they may be illuminative.

Final discharge destinations are identified as well as whether the discharge was delayed for at least 24 hours and the reason for such delays where they occurred (medical, social, and other). This provides important information that has been missing from all past studies of these issues.

3.4.9 ICD-9 Codes

The last section of the CARE tool documents the International Classification of Diseases (ICD)-9 codes associated with the patient's stay and submitted for payment.

3.5 Testing and Feedback During Development

The CARE tool and the items included in the CARE tool were extensively evaluated and tested during the development process and in specific reliability tests during the demonstration.

3.5.1 Pilot Tests

Two sets of pilot tests were conducted in the Chicago area. The first pilot test included only acute hospitals and LTCHs to test item appropriateness in these populations and to develop procedures that would complement current assessment and workflow practices. The second pilot test included all five types of providers and examined how well the tool worked in each setting and across a range of patients. The pilot tests ranged from 3 weeks to 6 weeks; settings with longer stay patients needed longer test periods to allow completion of both an admission and discharge assessment. The results of the pilot test were used to modify the CARE tool prior to publication in the July 2007 *Federal Register*.

Data collected in the pilot tests were tested for validity and reliability in each setting. Although the sample sizes were small in the pilot tests, they provided important preliminary information regarding the feasibility of using each item in the different treatment settings before testing the items in a national demonstration.

3.5.2 Stakeholder and Public Comment During CARE Development

As mentioned above, stakeholder and other public comments were incorporated in multiple stages and through multiple avenues. Provider associations were invited to ODFs to begin discussions regarding the appropriate domains and items to include in a uniform tool. The associations were also asked to nominate TEP members and to discuss which domains should be included in the tools. Nominations were received from a variety of sources including the Acute Long Term Hospital Association, American Association of Homes and Services for the Aging, American Health Care Association, American Hospital Association, American Medical Rehabilitation Providers Association, Commission on the Accreditation of Rehabilitation Facilities, Federation of Hospitals, Joint Commission on Accreditation of Healthcare Organizations, National Association for Home Care, National Association of Long Term Hospitals, and Visiting Nurse Associations of America. Participants included representatives from several large chains, including Amedysis, Genesis HealthCare, HCR ManorCare, HealthSouth Corporation, Hospital Corporation of America, Kindred Healthcare, and Select Medical Corporation, and individual providers and practitioners, including geriatricians from major teaching hospitals, such as Mayo. The second TEP was focused on gathering information from the research community representing measurement experts from each level of care to discuss the applicability and usefulness of specific measures. The input from these various sources was integrated with input from providers participating in two pilot tests.

Input was also sought from any person or group that wished to comment on the effort. Two ODFs were held in December 2006 and July 2007 to provide information on the demonstration and to invite input on the instrument's development. Additionally, RTI established and published an e-mail box, PAT-COMMENTS@RTI.ORG, to allow providers, clinicians, and other individuals to submit comments on the content of the tool and to bring to the team's attention issues that may be specific to one of their populations or settings that should be considered in designing this tool. These comments were incorporated in the clinical workgroup's efforts. Many of the national associations also published the address for submitting comments and invited their members to do so. Ongoing discussions with association executives over the past few years include those from the following organizations: NAHC, VNAA, AHCA, AAHSA, AMPRA, HealthSouth, Kindred, Select, NALTH, AHA, CHA, and Amedysis. Many

associations invited the project team to present information about the CARE items and the demonstration at their national meetings and at each of these presentations, attendees were invited to submit comments to the available Web site. Additional small group meetings were held by phone to discuss ideas regarding content or operational use of the tool in each level of care. Presentations at association meetings included, but were not limited to, special meetings or annual meetings of a variety of interested parties, including

- American Academy of Physical Medicine and Rehabilitation;
- American Association of Homes and Services for the Aging;
- American Congress of Rehabilitation Medicine;
- American Health Care Association;
- American Health Informatics Management Association: Long-Term Care Health Information Technology Summit;
- American Medical Rehabilitation Providers Association/National Institute on Disability and Rehabilitation Research: State of the Science in Rehabilitation Medicine;
- American Rehabilitation Nurses;
- California Hospital Association;
- National Association for Home Care;
- National Association for State Health Policy;
- National Association of Long Term Hospitals; and
- Uniform Data Systems.

Ongoing discussions with association executives over the past few years included those from the following organizations:

- Amedysis
- American Association of Homes and Services for the Aging
- American Health Care Association
- American Hospital Association
- American Medical Rehabilitation Providers Association

- California Hospital Association
- HealthSouth Corporation
- Kindred Healthcare
- National Association for Home Care
- National Association of Long Term Hospitals
- Select Medical Corporation
- Visiting Nurse Associations of America

The CARE tool was published twice in the *Federal Register* (July and November 2007) as part of the Office of Management and Budget Paperwork Reduction Act (OMB-PRA) review process. Each publication included a burden estimate based on the pilot test experience. These estimates ranged from a 30-minute assessment completion time for the healthier patient to 60 minutes in the LTCH or SNF where patients may be more complicated medically and/or functionally or have greater cognitive complications. These average times of completion reflect experience with the tool, following training on the appropriate measurement methods, and are consistent with current intake assessment times.⁵

During the OMB-PRA review cycles, comments were received from a wide range of the public, including clinicians, administrators, and others. Several issues were raised repeatedly by different types of respondents:

- *There was wide consensus and support for developing a standard assessment tool for use in the Medicare program.* Almost all respondents pointed to the importance of this effort for improving quality of care by standardizing the language used to measure illness and impairment and the value of having the Federal Government sponsor this work.
- *Respondent burden.* Participants were pleased with the relatively short length of this item set compared with the MDS or OASIS. Therapists in the SNFs and HHAs generally appreciated the CARE versions of the function items because they perceived them to better document patient impairment and improvement than the items in the current tools. Those working with pressure ulcers and wounds were pleased to have standard approaches suggested by the national wound organizations.
- *Suggestions were offered for item refinements, additions, and exclusions.* These suggestions were reviewed by the four RTI clinical workgroups, and a revised tool

⁵ These items are intended to replace non-uniform versions of the items already used and would not add any time relative to the current items. They added time in the demonstration because providers needed to continue collecting the mandated version for reimbursement while also collecting the test version during the study period.

was published in the October 31 *Federal Register* and used in the final PAC-PRD data collection.

3.5.3 Stakeholder and Public Comment During CARE Use

Presentations to the associations were also useful vehicles for inviting continued comments throughout the demonstration process. At each meeting, attendees were asked to visit the demonstration Web site (<http://www.pacdemo.rti.org>) to view the CARE tools and submit comments regarding the items' applicability to the Medicare populations treated in their setting. Feedback was requested on whether these items described differences in severity in their populations and whether any items were not applicable to certain populations or whether additional items were needed to distinguish among cases admitted to their setting with different treatment needs.

A comment section was included in the Web site that allowed respondents to identify the type of setting in which they worked, indicate their clinical licensure, and provide feedback on the items. This section remains active and continues to take comments from clinicians interested in participating in the item refinement process.

Additional comments also were requested on the last section of each assessment tool. As clinicians were completing the assessment, they were asked to provide feedback on the items in the tool as they applied to individual patients—for example, whether certain items were missing or alternatively, whether some were not relevant to the type of patient just assessed.

Last, every site that participated in the demonstration was asked to participate in an exit interview. This interview was designed to collect feedback on the process and the items used. Responses from these interviews are being incorporated into a “Lessons Learned” report that will complement the input from the reliability tests.

3.6 Reliability Study

An important question in deciding whether these standardized items should replace existing items in the Medicare payment systems is whether they are reliable when used with each of the PAC populations. This next section provides results from the reliability studies that were conducted in each of the five settings (acute, LTCH, IRF, SNF, and HHA). The results were used in selecting the final set of analytic items to include in the PAC-PRD models predicting resource intensity, readmission, and functional change.

Two types of reliability tests were conducted. The first, a traditional inter-rater reliability (IRR) study using paired assessments of patients, allowed analyses to focus on the reliability of the standardized items when applied to populations in settings other than those for whom the items were originally validated. The second type of test, where assessors in different settings rated uniform “hypothetical” patients, examined the degree of agreement when items were used by different disciplines in different settings. This second issue will be particularly important for considering patient-level differences as the beneficiary moves across an episode of care and is rated on the standardized health and function items in each setting.

Both sets of tests were conducted in a subset of participating PAC-PRD providers with a subset of clinicians who had already been trained on the standardized CARE items. Participants were retrained prior to the initiation of the reliability test to minimize effect differences due to time from training rather than item reliability.

3.6.1 Traditional Inter-Rater Reliability Testing

The first type of reliability test used a traditional IRR approach in which two raters of the same discipline each scored the same patient at approximately the same time. Staff from 27 providers participated in this test yielding 455 pairs of matched patient assessments. **Table 3-1** shows the number of providers participating and the number of paired assessments collected from each type of setting.

Table 3-1
IRR testing providers by type/level of care

Provider type	Number of providers enrolled	Paired assessment numbers
Acute hospitals	4	66 paired assessments
Home health agencies	8	102 paired assessments
Inpatient rehabilitation facilities	7	118 paired assessments
Long-term care hospitals	2	49 paired assessments
Skilled nursing facilities	6	121 paired assessments
Total	27	455 paired assessments

All acute, LTCH, IRF, and SNF facilities that participated in the IRR testing were asked to complete 15 to 20 paired duplicate assessments, and HHAs were asked to complete 10 to 15 duplicate assessments. Facilities were asked to identify a set number of fee-for-service Medicare patients for inclusion in the testing, representing a range of function and acuity. For these identified patients, providers were instructed to have pairs of raters complete both patient assessments at the same time upon admission or, at a minimum, within the 48-hour reference window. Patients were assessed by staff pairs matched by discipline (e.g., two nurses, two physical therapists).

Responses were obtained by one or more of the following predetermined, matched methods: direct observation of the patient (includes hands-on assistance), patient interviews (with each team member taking turns conducting and observing patient interviews), interviews with relatives/caregiver of the patient for certain items, and interviews with staff caring for the patient and/or chart review. Rater pairs were instructed to determine in advance which methods would be used to score the particular CARE tool items and to have both raters use the same methods. Raters were encouraged to divide hands-on assistance to the patient as evenly as possible for CARE items that required hands-on assistance, such as the functional status item “sit to stand.” For patient interview items, such as those in the Temporal Orientation/Mental Status, Mood, and Pain sections, raters were instructed that one rater could conduct the entire interview, or the raters could alternate questioning. Raters were instructed not to discuss CARE item

scoring during the CARE assessment, nor to share item scores until the data were entered into the CMS database and finalized.

Item Selection for Testing—CARE tool items selected for IRR testing fell into one (or more) of the following categories: items that are subjective in nature, items that have not previously appeared in CMS tools (i.e., new CARE items), items that influence payments or are used in payment models currently, or items not previously tested in certain settings. Items excluded from the reliability tests included less subjective items such as ICD-9 codes and the use of major treatments (yes/no indicators based on medical charts and patient observation for resources such as ventilators, hemodialysis, and central lines).

Analytic Methods—RTI used two analytic approaches for assessing the IRR of the CARE tool items, following closely the methods used in prior CMS assessment IRR analyses. For continuous items, RTI calculated Pearson correlation coefficients to show the extent of correlation between two raters on the same item. For categorical items, RTI calculated kappa statistics, which indicate the level of agreement between raters using ordinal data, taking into account the role of chance agreement. Acceptable levels of agreement are typically moderate or better. **Table 3-2** shows the ranges commonly used to judge reliability based on the kappa results.

Table 3-2
Standard interpretation of kappa scores

Agreement	Kappa
Poor agreement:	0
Slight agreement:	0.01–0.20
Fair agreement:	0.21–0.40
Moderate agreement:	0.41–0.60
Substantial agreement:	0.61–0.80
Almost perfect agreement:	0.81–1

Both weighted and unweighted kappas are reported; the two approaches make different assumptions about the data. Unweighted kappa assumes the same “distance” between every one unit difference in response across an ordinal scale (e.g., for the CARE functional item scale range 1 to 6, an unweighted kappa assumes the difference in functional ability between a score of 1 = dependent and 2 = substantial/maximal assist is the same as the difference in functional ability between 5 = setup or clean-up assistance and 6 = independent). Weighted kappas can be calculated to assign different distances between responses. Standard Fleiss-Cohen weights, or quadratic weights, which approximate the intraclass correlation coefficient and are commonly used for calculating weighted kappa, were used in this analysis to allow comparison with prior analyses. This strategy puts lower emphasis on disagreements between responses that fall “near” to each other on an item scale. Weighted kappas using Fleiss-Cohen weights are influenced by the number of response levels in a scale and tend to be higher when there are more levels available. Kappas, weighted or unweighted, can be influenced by the prevalence of the outcome or characteristic being measured. If the outcome or characteristic is either very rare or very

common, the kappa will tend to be low because kappa attributes the majority of agreement among raters in these instances to chance. Kappa is also influenced by bias, and, if the effective sample size is small, variation may also play a role in the results. We report both weighted and unweighted kappas to give the range of agreement found under the two sets of assumptions. RTI also calculated a separate set of kappa statistics (unweighted and weighted where applicable) for items that excluded the nonordinal (or letter code responses) by setting these items to missing. These results show the reliability for items for cases that were coded and exclude cases with missing data.

3.6.2 Results

Overall, the results showed very good agreement on most items. Across all 146 items tested, only 17 percent had a rating lower than 0.60, including both the unweighted and weighted items and samples with and without letter codes included. Looking just at the weighted kappas for samples that exclude letter codes or unweighted kappas where appropriate, 13 percent (19 items) of the 146 items had a reliability of 0.70 or lower. Items with poorer agreement among any of the samples (less than 0.60) tended to be items with fewer responses (e.g., items where the response code was “other” or “tube feeding” and “comatose,” for which few cases were included). However, a few items with reasonable sample sizes also appeared to be less reliable, such as certain components of the swallowing item (“complaints of difficulty or pain when swallowing,” “holding food or liquid,” and “loss of liquid when swallowing”). These lower reliability ratings were offset in the swallowing item by less discretionary components, such as “no intake by mouth” (NPO; 0.97) and “no impairments” (0.84). Other poor-scoring items included “walking 150 feet,” “light shopping,” and “laundry.”

Agreement was fairly high across providers on most items, with some variation across the different domains. These are discussed in more detail below.

Prior Function—“Prior functioning” had high rater agreement, with codes on each item ranging from 0.75 to 0.86. “History of falls” also had very high agreement between raters (0.88). These kappas were fairly consistent across the five types of providers, although IRFs tended to have lower agreement on this interview item (0.50 for weighted and 0.54 for unweighted self-care). HHAs had the second lowest ratings (between 0.74 and 0.70), and each of the other providers had even higher rates of agreement on this interview/history item.

Skin Integrity—All kappas for the evaluated pressure ulcer items indicated substantial or near-perfect consistency. The lowest weighted kappa was for the “unstageable ulcer” (0.68); the rest of the pressure ulcer items ranged from 0.70 to 0.83. The major wound items also had substantial or almost perfect ratings, ranging from 0.64 for agreement on “delayed healing” to 0.93 for agreement on “vascular ulcers.”

The intact turning surfaces item was less reliable, with results ranging from 0.21 for “other surfaces not intact” to 0.76 for “back/buttocks not intact.” The two items with potential usefulness in this group are “back/buttocks not intact” (0.76) and “skin surfaces for all turning surfaces is intact,” which also had substantial agreement (0.66).

Looking across settings, agreement was almost perfect for the pressure ulcer item 3.G2 (“Does this patient have one or more unhealed pressure ulcer[s] at stage 2 or higher or unstageable”), with kappas for HHAs, LTCHs, and SNFs each indicating almost perfect agreement (0.82 to 0.92). Kappas for acute hospitals demonstrated substantial agreement (0.73), while inter-rater reliability in IRFs indicated moderate agreement (0.58). This result may be due to the wider range of disciplines that may assess pressure ulcers in IRFs. For CARE tool item 3.G6a (“Skin for all turning surfaces is intact”), LTCHs exhibited almost perfect consensus between raters (0.87), while kappas for both acute care providers and HHAs indicated substantial agreement (0.64 and 0.72, respectively).

Cognitive Items—The Brief Interview for Mental Status (BIMS) items were taken from the MDS 3.0 cognitive section and had very strong agreement, with weighted kappas ranging from 0.71 to 0.91 and unweighted kappas ranging from 0.62 to 0.86. This agreement held true across all providers in looking at the “knows year” item, with the lowest scores in SNFs (0.73) and the highest scores in IRFs (1.0). The kappas were highest for the temporal orientation items (4.B3b) at 0.86 and above and “recall of three words” (4.B3c) at 0.89 or above for the second recall item. The first memory item, “repetition of three words,” was slightly lower but still substantial at a kappa of 0.71.

The CAM measure, used to measure delirium in hospital patients, had substantial agreement for “inattention” and “disorganized thinking” (0.70 to 0.73); however, “altered level of consciousness” and “psychomotor retardation” were lower at 0.58 and 0.48, respectively. Across providers on the inattention item (4.D1), IRFs had the highest agreement at 0.82 for the weighted kappa and 0.74 for the unweighted kappa. The rest of the providers’ rates of agreement were all above 0.60. This result is consistent with the existing literature on this item, which ranges from 0.59 to 0.82.

Depression/Sadness Items—The CARE included two depression items: the PHQ-2 and the PROMIS item. The PROMIS item was based on the SF-36, which was developed for the general population, including the healthy population. The kappas suggest that the PHQ-2 items were slightly more reliable across the acute and PAC populations than the “feeling sad” item (more kappas above 80, although the lowest kappa on the “feeling sad” item was 0.742), suggesting both are fairly reliable in these populations. For the PHQ-2 item 4.F2c (“feeling down, depressed, or hopeless”), kappas with “unable to answer” or “no response” excluded indicated almost perfect agreement, with values ranging from 0.81 to 0.89 for all provider types except acute hospitals, which did not have this item on their tool.

Pain Items—The interview-based pain items (4.G1 through 4.G5) had substantial to almost perfect kappas whether or not coded nonresponse items were included in calculations (weighted kappa range: 0.79 to 0.88). Kappas on the “pain presence during the last 2 days” (4.G2) item indicated almost perfect agreement (ranging from 0.88 to 0.94) in all care settings except for SNFs, where kappa values indicated substantial agreement (0.72).

Observational assessment pain items had lower kappa values than the interview items, as expected, but were still substantial for “nonverbal sounds,” “vocal complaints of pain,” and “facial expressions” (range 0.61 to 0.66). “Protective body movements or postures” (4.G6d) had a lower kappa at 0.42.

Impairment Items—The bowel and bladder items showed substantial consistency between raters, with kappas ranging from 0.60 to 0.90, with most items over 0.70. Kappas appeared to be a bit higher for bladder items, although bowel management kappas may have been affected by lower prevalence of impairments in bowel management. The lowest weighted kappas for bladder incontinence were in LTCHs (0.66).

Swallowing signs and symptoms had more variation in scores, with high agreement for “NPO: intake not by mouth” (5.B1e) at 0.97 but offset by “complaints of difficulty or pain with swallowing,” which had the lowest score in this group at 0.46. “Holding food in mouth” and “loss of liquids” had scores of 0.56 and 0.57, respectively. “Coughing or choking” and “other signs and symptoms” had substantial agreement, and raters were almost perfect when evaluating if a patient had “no signs or symptoms” (0.84). Across providers, the lowest agreement on this item was found in HHAs and LTCHs, which had kappas of 0.64 and 0.67, respectively.

The hearing, vision, and communication comprehension items on the CARE tool include four items taken from the MDS 3.0. The goal of these items is to identify the level of impairment as mild or moderately impaired, severely impaired, or not impaired. The kappa statistics for these were all strong, with weighted kappas between 0.74 on sight to 0.80 on hearing.

Both the weight-bearing and grip strength items showed kappas above 0.71, although the scores varied by individual items. The weight-bearing items ranged from 0.71 for agreement on upper right extremity to 0.90 for agreement on lower left extremity. Agreement for grip strength ranged from 0.75 in the left hand to 0.85 in the right hand.

Respiratory status also had very high kappas, with weighted kappas ranging from 0.79 to 0.87 for items with and without oxygen, respectively.

Kappas for endurance items, both mobility and sitting items, showed substantial agreement, whether weighted or unweighted (0.69 to 0.76 or 0.62 to 0.71, respectively). For the “sitting endurance” item (5.G1b), acute hospitals and SNFs had the highest kappas (0.78 and 0.75), respectively, followed by HHAs (0.74). IRFs had the lowest agreement at 0.41 for the weighted kappas.

Functional Status—The CARE tool includes a core set of six self-care items and five functional mobility items that are scored on all patients. Items represent a range of difficulty. Many of these are modified from existing items on the OASIS, MDS 3.0, and IRF-PAI.

Kappa statistics for all core items, self-care, and mobility indicated substantial agreement among raters, with weighted kappa at 0.78 or above for the overall sample. The unweighted kappas were slightly lower, ranging in the mid-60s, with the exception of the tube feeding and oral hygiene items, which were lower (0.22 and 0.59, respectively). (Tube feeding scores were low because of low prevalence of tube feeding in our sample population.) The weighted kappa values remained consistently high across provider type, with a few exceptions. Agreement in the eating score was lower for HHAs (0.61); the oral hygiene and chair transfer scores were lower for LTCHs at 0.55 and 0.52, respectively.

Mobility items also had high agreement scores, ranging from 0.56 for “walking 150 feet” (which had small numbers) to 0.90 for “transfers” in the weighted scores. Unweighted kappas were slightly lower, ranging from 0.68 for “toilet transfer” to 0.76 for “sit to stand.” These relatively high levels of agreement were consistent across all five settings, with weighted kappas for “lying to sitting on side of bed” ranging from 0.72 for LTCH cases to 0.87 for SNF cases. For “sit to stand” items, agreement ranged above 0.81 (LTCHs were excluded for small numbers). “Chair/bed transfers” were also consistently high across providers, with the lowest scores being 0.78 in IRFs to the highest of 0.93 in SNFs.

Supplemental self-care items also scored consistently high, with each weighted kappa being above 0.8 and the unweighted kappas consistently ranging between 0.63 (“shower/bathe self” or “wash upper body”) and 0.74 (“picking up object”). Similarly, supplemental mobility items had kappas of 0.80 or above for weighted kappas and 0.64 (one-step curb) to 0.78 (“walk 10 feet on uneven surface”). Again, there was slight variation across providers, but all weighted kappas ranged above 0.70, with the one exception of “rolling left to right” in LTCHs, which showed kappas of 0.52.

IADLs all had weighted kappas of 0.7 or above except for light shopping and laundry (0.52 and 0.48, respectively). Notably, these items applied to many fewer cases due to medical complexity or the inability of staff to observe the patient’s performance of this type of activity in these settings. This finding was particularly true for medication management in the inpatient setting.

Overall Plan of Care and Health Status—Overall plan of care items including the overall health status item were also examined. The two plan-of-care items had reasonable kappas of 0.82 or 0.76, but the patient’s overall status had lower kappa scores (0.68 for weighted and 0.59 for unweighted). At the provider level, there was variation by type of provider. Acute hospitals, HHAs, and LTCHs had kappas of 0.67, 0.73, and 0.74, respectively, while IRFs had kappas of 0.35 and SNFs of 0.57.

Summary of IRR Tests—These results suggest that most of the standardized versions of the assessment items have strong reliability within and across settings. This finding is not unexpected, given that most of the CARE items are standardized versions of health status concepts already being measured in each setting. A few items had lower reliability, suggesting that their use across settings without greater development may be limited. Items with lower reliability include the skin integrity item measuring the components of turning surfaces not intact; the observational pain item measuring pain based on protective body movement or postures; several components of the swallowing items, such as complaints of difficulty, holding food in cheeks, and loss of liquids when eating/drinking; and the three IADL items of light shopping, laundry, and public transportation.

All other items scored reasonable levels of reliability. Differences across settings were present, but each setting still had acceptable levels of reliability within the setting, suggesting that these items could be used to measure a patient’s progress in a standardized way across an episode of care.

3.7 Reliability Testing of Clinician Agreement Across Settings

A limitation of the within-facility IRR approach is that the expected agreement across settings is unknown. Therefore, we conducted video-based case studies to test agreement across sites, type of providers, and clinicians. Nine videos were developed to present a standardized set of information to clinicians in each of the five settings. The videos varied in the severity of the patient presented and the specific clinical, cognitive, and functional profile shown. Participating providers were randomly assigned to watch one or two of the videos and use the information presented on the videos to complete CARE tool items. Two analytic approaches were used for assessing the video reliability of the CARE tool items. The approaches were consistent with the methods used by Fricke et al. (1992) to assess the reliability of the FIM items using videos. First, for each CARE item included in at least one of the nine videos, percent agreement was calculated with the mode response for the full sample. Unlike the approach used by Fricke and colleagues, RTI did not consider agreement at one response level above and below the mode; instead we used a stricter approach looking at direct agreement only. In the second approach, percent agreement with the internal clinical team’s consensus response was also calculated. This second measure not only gives an indication of item reliability, but also reflects on training consistency. These results are very conservative estimates of reliability because they are not restricted to responses by those clinicians in the sample who typically score a domain.

Table 3-3 shows the number of providers and assessments collected in each setting. Of the 550 assessments collected, 47 percent were completed by registered nurses (RNs), 21 percent by physical therapists, 14 percent by occupational therapists, 8 percent by “other” (largely licensed practical nurses [LPNs]), 6 percent by case managers, and 5 percent by speech language pathologists.

Table 3-3
Video reliability testing providers by type/level of care

Provider type	Number of providers enrolled	Assessment count
Acute hospitals	3	15
Home health agencies	9	118
Inpatient rehabilitation facilities	8	237
Long-term care hospitals	3	114
Skilled nursing facilities	5	66
Total	28	550

In general, the results showed substantial agreement among the disciplines; for most items and disciplines completing assessments, agreement with the mode or the internal clinical team was 70 percent or higher. The variation here is generally within the higher levels of agreement. These results are not surprising in that most clinicians have to address the types of items measured in the tests and are therefore familiar with evaluating the patient for these types of items. This type of reliability test is useful for understanding the extent to which clinical background may result in a different scoring of the patient’s health status.

3.7.1 Prior Functioning

Rates of agreement for all the prior functioning items were all above 0.69. In general, nurses, including both case managers and “other” (LPNs), scored lower on agreement for the prior functioning measures than did the physical or occupational therapists. Differences were within 5 to 10 points of each other, depending on the items. This finding was true in both the comparisons with the modal responses and the expert clinical team responses.

3.7.2 Skin Integrity

Results for the pressure ulcer items demonstrated particularly high agreement, with the lowest proportion being 0.5 for the speech pathologists identifying stage 3 ulcers relative to the mode. This finding is not surprising, because this item is generally not one that a speech pathologist would evaluate. Physical therapists had the highest agreement with the mode for identifying risk of pressure ulcer (0.94) or presence of a stage 2 or greater ulcer (0.98), followed by RNs, with a modal agreement of 0.88 and 0.95, respectively.

3.7.3 Cognitive Status, Mood, and Pain

Results for the cognitive status and mood items showed very high levels of agreement with the mode and clinical team, rarely falling below 90 percent. The minor exception to this trend was item IV.C., “observation of cognitive status” (C1), which is used when the Brief Interview for Mental Status (BIMS) cannot be administered. For this item, levels of agreement showed a great deal of variability among disciplines, varying from 40 percent among physical therapists to 76 percent among RNs and 100 percent for case managers. However, it is important to recall that because the standard method of assessing cognitive status on the CARE tool is the BIMS, the observation of cognitive status item was only used on one of the nine videos (Video 9). Among RNs, who were the largest group assessing this particular video ($n = 37$, or 51 percent), a substantial level of agreement was observed (76 percent).

Pain items also showed fairly high levels of agreement, although speech therapists had lower levels of agreement (0.70) for identifying pain, while occupational therapists (0.92) and physical therapists (0.91) had the highest rates of agreement, followed by RNs (0.84).

3.7.4 Impairments

The bowel and bladder items showed substantial agreement with the sample mode and clinical team response, with most items over 80 percent among all disciplines. In general, slightly lower levels of agreement were observed among clinicians who self-reported as “other,” although agreement levels were still moderate to substantial even in this group of clinicians. The item for “frequency of bladder incontinence” (A3a) had slightly lower levels of agreement than the other bladder and bowel items, with speech therapists having the lowest level of agreement (0.50). Again, it is important to note that these items are not usually evaluated by this type of clinician.

Swallowing signs and symptoms also showed substantial agreement among raters (generally 80 percent or above), with the category of “other” exhibiting slightly lower levels of agreement. Speech pathologists had the highest levels of agreement (0.92) on the “usual

swallowing ability” item. Results were more mixed on the “signs of swallowing disorder” item, which also had lower IRR on several components.

Hearing, vision, and communication items all had fairly high rates of agreement across disciplines, with the “other” category (LPNs, mostly) scoring the lowest levels of agreement, followed by RNs for “understanding content” and “ability to hear”; the proportion agreeing were 0.81 and 0.88, respectively. Speech pathology tended to have the highest rates of agreement with the mode and internal clinical team on these items, followed frequently by physical therapists or occupational therapists.

Respiratory status had variable rates of agreement depending on whether the patient used oxygen. Presence of any respiratory impairment had the highest rates of agreement for occupational therapists, RNs, and speech pathologists (0.93, 0.87, and 0.94, respectively). When rating the level of exertion with oxygen when a patient becomes dyspneic, speech pathologists and occupational therapists had the highest rates of agreement (0.73, 0.75), compared with raters in other disciplines (with rates between 0.48 and 0.56). This item had eight potential responses, so it is not surprising that the rates of agreement were lower, given our strict counting of exact agreements only.

Endurance items, both sitting and mobility, had relatively high levels of agreement across the core screening item (88 to 100 percent), whereas the supplemental items showed more variation, with speech pathologists having the lowest levels of agreement (0.75) and case managers and physical therapists having the highest rates of agreement.

3.7.5 Functional Status

The core functional status items also showed high levels of agreement with the mode and clinical team for all items, typically greater than 70 percent. The notable exception to this trend was among the clinicians self-reporting their discipline as “other”; they consistently had the lowest levels of agreement among all core self-care items, ranging from 0.50 to 0.72 percent agreement.

Supplemental self-care items such as the ability to “wash, rinse, and dry the upper body” and to “bathe self in the shower or tub” and mobility items such as “rolling from lying on the back to left and right side,” “move from sitting on side of the bed to lying flat on the bed,” “bend/stoop from a standing position to pick up a small object from the floor,” and “put on and take off socks and shoes or other footwear” suggest a fair amount of variability between disciplines. For the self-care items, the occupational therapists, physical therapists, and RNs reported substantial levels of agreement with both the mode and clinical team, ranging from 65 to 94 percent. Case managers, speech therapists, and the “other” category tended to show slightly lower levels of agreement on certain items (e.g., 50 percent for “other” and 63 percent for speech therapists on “shower/bathe” and 50 percent for case managers on “picking up an object.”)

Similar trends were observed on supplemental function items (C7a–h) and the majority of the IADLs (items C8–C16). For items C7a–h, agreement with the mode and the clinical team response generally ranged from 70 to 100 percent, although case managers and the “other” discipline category reported suboptimal agreement on some items.

For the IADL items (C8–C16), agreement with the mode was generally substantial (exceeding 75 percent), although several items had more moderate levels of agreement overall. These items were “medication oral,” “medication mist,” “wipe down surface,” and “laundry” (C10, C11, C14, and C16). Among occupational therapists, physical therapists, and RNs, agreement for these items tended to fall in the more moderate range of 50 to 72 percent, with agreement among speech therapists, case managers, and the “other” category often significantly lower.

These analyses are useful for examining the reliability of these items across settings, disciplines, and training experiences. These video-based assessments show that when presented with a standardized interview or observation, the clinicians were able to apply the item definitions consistently. Although this approach differs from clinical practice, where assessment and interview techniques may vary, it is consistent with the approach used in FIM-credentialing examinations (Fricke et al., 1992). Item reliability is a difficult area to measure, but the results suggest that it remains consistently high across disciplines, with some variation as expected in specific items. These results are useful for considering cross-setting measurement constraints.

3.8 Summary of the Results

Overall, the standardized CARE items are reliable items when used across settings and by different disciplines. The levels of agreement varied, but most were above 0.70; a few items appeared weaker, such as certain aspects of swallowing measurement, walking 150 feet, light shopping, and laundry.

Levels of agreement varied minimally across disciplines, suggesting that the definitions of the items were clear and could be used consistently with proper training. The reliability statistics were mostly consistent with past application of these items to one population or another. The tests were also useful for identifying the few items that had lower kappa statistics, such as laundry, which could be eliminated from use in the analytic models. It is not surprising that most of the items were reliable when applied in different settings because, in general, they represent concepts already measured in each of the different sites. Extensive training and help desk assistance were provided throughout the demonstration, which likely increased clinicians’ skills with these items.

3.9 Next Steps: Use of a Flexible Electronic Standards-Based Instrument

Section 4 of this report discusses specifics regarding the collection of data for the purposes of this demonstration. One of the features discussed briefly is the use of a secure Internet-based application that allowed authorized persons at the participating providers to submit CARE assessments directly to the CMS data centers. The electronic collection system also allowed participating providers to upload data obtained by their electronic records for those items where the exact definition of the items matched what was collected by CARE.

The impetus behind this effort was, in part, to move toward electronic standardization in addition to the other ways that the development of the CARE tool sought to standardize the assessment of patients between provider types. CMS’ vision developing CARE was to move from multiple incompatible assessment instruments to one standardized set of clinically relevant data that applies federally and nationally recognized health information technology (HIT)

standards. Use of broadly adopted HIT standards will allow for the safe, secure, electronic exchange of critical health information among authorized users. CARE data, as shown by the PAC-PRD, can be collected on paper or through an electronic platform.

The CARE data set was designed as a dynamic set of items that could be drawn from a “library” or registry of standards-based items for measuring the different concepts. This first generation of CARE targets basic core and supplemental items for measuring frequently occurring conditions in the Medicare populations, such as medical, surgical, and functional conditions. In the future, standardized subsets of CARE data or modules that are more specific to a particular condition and/or provider setting could be drawn from the registry storing the standardized CARE library of elements and concepts. This approach will allow items to be used with other data commonly collected for care planning and allow item modules to be added in the future as more of the clinical items used in quality monitoring and survey and certification become integrated. For example, additional standards-based items could be added to capture individual patient preferences for care treatments and items that measure the degree to which individual’s preferences and goals have been met. Effectively, CARE has been designed to evolve over time to incorporate a broader range of items to address patient-centered care planning, quality measurement and reporting, and other emerging needs. It has been designed to meet Federal IT requirements for standards-based exchange of meaningful health information among authorized users. A variety of efforts, separate from the work done under this contract, have been undertaken to develop and test electronic specifications of the CARE items.

3.10 Summary of the Section

This section described the creation and preliminary testing of CARE, a standardized set of assessment items for use at discharge from acute hospitals and at admission and discharge from PAC settings. In the creation of CARE, CMS achieved its goal to develop a set of standardized assessment items that are useful, clinically relevant, and grounded in scientific evidence and stakeholder input. CARE lays the groundwork for using a standardized data set across all providers to assess beneficiaries’ progress and outcomes achieved in relation to resources used in various health care provider settings. CARE successfully meets the legislative directive to collect data predictive of outcomes and resource utilization to guide quality and payment policy development. Additionally CARE provides a standardized data collection vehicle for measuring beneficiaries’ health and functional status longitudinally across episodes of care. Analysis of the data collected using CARE will be presented in subsequent sections of this report.

Although still preliminary, the reliability and validity work performed on the CARE items was extremely promising and equivalent to other mandated assessment instruments (see Gage et al., 2008). The standardized CARE items are reliable items when used across settings and by different disciplines. The levels of agreement varied, but most were above 0.70; a few items appeared weaker such as certain aspects of swallowing measurement, walking 150 feet, light shopping, and laundry. Levels of agreement varied minimally across disciplines suggesting the definitions of the items were clear and could be used consistently with proper training and documentation. The reliability statistics were mostly consistent with past application of these items to one population or another. The tests were also useful for identifying the few items that had lower kappa statistics, such as laundry, which could be eliminated from use in the analytic

models. It is not surprising that most of the items were reliable when applied in different settings because, in general, they represent concepts already measured in each of the different sites. Extensive training and help desk assistance were provided throughout the demonstration, which likely increased clinicians' skills with these items.

CARE successfully moves CMS and providers forward from the use of multiple incompatible assessment instruments to one standardized set of clinically relevant data that applies federally and nationally recognized HIT standards. Use of broadly adopted HIT standards will allow for the safe, secure, electronic exchange of critical health information among authorized users. The organization of the tool and the work performed related to the creation of an Internet-based data collection application and import process have created an electronic system that is flexible for easy, rapid accommodation of future clinical and technological advances and electronically based on federally and nationally recognized standards for interoperability across settings. Both the tool and the electronic underpinnings were created with extensive input from and support by stakeholders.

SECTION 4 DEMONSTRATION METHODS AND DATA COLLECTION

This section discusses the general data collection approach used in the Post-Acute Care Payment Reform Demonstration (PAC-PRD) and addresses issues of the representativeness of the sample. Two types of data were collected in five types of participating providers: general acute care hospitals, long-term care hospitals (LTCHs), inpatient rehabilitation facilities (IRFs), skilled nursing facilities (SNFs), and home health agencies (HHAs). First, all five types of providers collected the standardized Continuity Assessment Record and Evaluation (CARE) assessment items discussed in Section 3 in both the initial and the supplemental phases of data collection. These data provided standardized measures of patient acuity for each of the enrolled beneficiaries. Second, cost and resource utilization (CRU) data, which provided staff-time measures for treating each of the enrolled beneficiaries, were collected in the four types of PAC providers during both initial and supplemental data collection phases and also in acute providers during the supplemental phase of data collection. Below, we describe market area and provider selection, provide data on PAC use in the participating market areas to demonstrate the range of PAC use patterns captured in this study, and provide an overview of data collection processes.

4.1 Sample Framework

The data collection presented in this report comes from two phases of collection. The initial phase of collection was the one designed to address the primary issues raised by the initiating language of the Deficit Reduction Act. The initial phase of collection featured a hierarchical clustered design: patients within facilities, facilities within markets, and markets within the United States. The approach taken was to create a “snapshot” of patient acuity and patient resources within each participating provider. This snapshot approach provides information on the type of patient treated in each setting in proportion to how often they are seen. Thus, patients at different severity levels are in the sample, as are patients at different points in their care trajectory (at an acute hospital, in their first or subsequent PAC setting, in PAC after a readmission, etc.).

The supplemental phase of data collection had the goal of gaining insight into providers that treat medically complex patients and providers who also treat patients commonly treated in IRFs. Because of a shift in analytic focus, the sampling methods of this supplemental phase differed from the initial phase in what market and provider characteristics were targeted. Where appropriate, the analysis presented may include data from the initial collection or a combination of the initial and supplemental collections.

4.1.1 Market Area Characteristics

In selecting the market areas and participating sites for the initial phase of data collection, we attempted to account for the following characteristics:

- beneficiary/patient representativeness, particularly focusing on variations in patient types or primary conditions treated and targeting those most likely to be treated in substitute settings, but also capturing typical conditions treated in the Medicare PAC populations

- market variation in terms of the types of PAC settings available in each area, including variation in availability of hospital-based compared with freestanding providers
- practice variation in terms of the types of patients admitted to different PAC settings, including variation in severity of condition and nature and intensity of treatments
- geographic variation, both regionally and by urban, suburban, or rural populations served

In examining the results of the analysis, it is important to keep in mind that the sample chosen was meant to represent illustrative types of experiences and was not meant to represent the Nation as a whole without weighting. For example, the acute hospital units selected to participate have sicker patients and patients more likely to go on to PAC services than the “average” hospital because these types of units were deliberately chosen for analytic reasons. Similarly, the sample oversamples rural areas, LTCHs, and IRFs in order to have sufficient sample sizes to address issues within those populations. The supplemental phase of data collection focused on markets differing on the availability of LTCHs or IRFs to provide data on patients of interest treated in LTCHs or IRFs or in other types of providers in areas of the country with few or no LTCHs or IRFs.

4.1.2 Market Area Selection

Figure 4-1 is a map showing the geographic distribution of the market areas selected for the initial data collection phase of the PAC-PRD. The market areas include providers located within a 2-hour driving distance from the center of each city listed below. This distance was chosen for the practical purposes of holding a “train-the-trainer” session at one location in each market area for all participating providers and allowing RTI staff to conduct site visits at all participating providers at the start of data collection. All providers in these geographic areas meeting minimum quality-of-care thresholds were considered for recruitment. Quality of care was a consideration for participation in the study to ensure that policy discussions and decisions were not based on data from poor-performing providers. Hospital Compare, Nursing Home Compare, and Home Health Compare were the source of quality information on providers. The market areas are centered around the following areas:

- Boston, Massachusetts
- Chicago, Illinois
- Columbia, Missouri
- Dallas, Texas
- Lincoln, Nebraska
- Louisville, Kentucky

- Omaha, Nebraska
- Portland, Oregon
- Rochester, New York
- Sacramento, California
- San Francisco, California
- Seattle, Washington
- Sioux Falls, South Dakota
- Tampa/Lakeland, Florida
- Wilmington, North Carolina

These markets were chosen to represent differing geographic areas, urban or rural status, and supply of PAC providers. For example, Dallas and Boston were chosen for their high provider supply—in particular, the high number of LTCHs and freestanding IRFs available. Rochester was chosen because of the absence of LTCHs and freestanding IRFs. Each of the selected markets provides important information on PAC utilization by provider supply. **Table 4-1** summarizes the number and type of providers available in each of the areas targeted for recruitment and other market characteristics. For the purposes of this discussion, the Sacramento and the San Francisco radiuses are combined into one market referred to as San Francisco. Similarly, the Portland and Seattle 2-hour radiuses are collectively referred to as Seattle, and Omaha and Lincoln are collectively referred to as Lincoln.

The supplemental markets were chosen to represent providers that treat high-acuity patients. Thus a primary focus for selecting markets was to select areas that had either a strong LTCH presence or ones that had low access to LTCHs that were not captured in the initial phase of data collection. Unlike in the initial data collection phase, the focus of the supplemental data collection phase was on targeting a few providers within a market area. The markets contained both single and multiple providers in a geographic region. The supplemental provider markets are listed below and illustrated in Figure 4-1.

- Baltimore, Maryland
- Cleveland, Ohio
- Detroit, Michigan
- Lynchburg/Roanoke, Virginia
- Los Angeles, California

- Manchester, New Hampshire
- Norfolk, Virginia
- New York, New York
- Philadelphia, Pennsylvania
- Portland, Maine
- Raleigh, North Carolina

4.2 PAC Utilization and Episodes of Care in Initial PAC-PRD Market Areas

This section presents the results of analyses of claims data, looking at the patterns of PAC use in each of the initial data collection market areas to demonstrate the range of service use patterns captured in the PAC-PRD markets. These data are based on analyses of 100 percent of acute hospital discharges in each of the initial data collection PAC-PRD market areas in 2006. Note that these data were not examined for the supplemental market areas. The data are intended to provide additional background information for the demonstration analyses by providing market-level descriptive statistics on PAC utilization. As mentioned in Section 4.1, the goal of the sample selection was to be able to examine populations of interest; the sample was not meant to be representative as a whole without weighting. Thus, variation in practice patterns between markets was a desired and sought-after characteristic. By including the range of observed service use patterns, we can feel confident that the sample reflects utilization patterns observed across the country, given different supply and practice patterns. The analytic file development and market area analyses in this section were supported by the Assistant Secretary for Planning and Evaluation (ASPE) in collaboration with CMS.

4.2.1 Use of PAC Services and Use of Specific Provider Types in Participating Markets

RTI built an episode file to examine patterns of PAC use in each of the initial data collection market areas. In building this file, markets were defined as providers in ZIP codes within a 2-hour driving distance of each of the selected cities, consistent with the market definition outlined above. Episodes were defined as starting with an index hospitalization and ending with the last hospital or PAC claim before a 60-day gap in acute and PAC service use. This variable-length episode definition was used in earlier work with ASPE (Gage et al., 2009a) and is consistent with the Medicare spell of illness. Although more recent work with ASPE has examined shorter episode definitions, this longer, 60-day episode definition allows for an understanding of the complete pattern of service use associated with an index event—an acute care hospitalization. Index acute care hospitalizations were selected for the episode file if they occurred after a 60-day period without acute or PAC service use.

Tables 4-2 through **4-4** describe the use of PAC nationally and in each of the initial data collection market areas to demonstrate the range in PAC service use reflected in the PAC-PRD market areas. Table 4-2 demonstrates how the market areas selected for the PAC-PRD vary in

the percentage of beneficiaries discharged to PAC and in the types of services used after discharge from acute hospitals. The results of this analysis demonstrate the range in beneficiary use of PAC services in lower use areas such as Sioux Falls (30.9 percent) compared with higher use areas such as Boston (47.8 percent). These results also demonstrate the variation in the types of PAC services used across the market areas. For example, in Dallas, a high proportion of beneficiaries were discharged to LTCHs (8.0 percent) and IRFs (20.8 percent), compared with Seattle (LTCH: 0.3 percent, IRF: 6.7 percent) and Rochester (LTCH: 0 percent, IRF: 4.6 percent), where there was little use of LTCHs, relatively low use of IRFs, and high use of SNFs and HHAs as a first site of PAC.

Tables 4-3, 4-4a, and 4-4b show the differences in PAC episode length of stay and payments for episodes overall and by service type in the participating market areas. Beneficiaries discharged to PAC in Dallas had the longest episodes at 98.3 days, compared with the shortest episodes observed in Sioux Falls at 48.9 days. The high use of HHA services in Dallas likely influenced the long episode length of stay. For this analysis, an episode of care might include multiple HHA episodes, as was often the case in the Dallas market. In Dallas, 68.6 percent of beneficiaries using PAC services had at least one claim for HHA, with a mean of 48 visits per episode of care. This mean number of visits per stay in home health is twice that of the market area with the second highest mean number of visits, Boston, at 23.4 visits per episode of care. The percentage of beneficiaries with readmissions during their PAC episodes also varied significantly across market areas. In Chicago, 29.2 percent of beneficiaries had a readmission during their episode, compared with less than 23 percent of beneficiaries in Sioux Falls, Portland, Sacramento, and San Francisco. These data confirm that the market areas selected for the demonstration are reflective of a wide range of PAC supply and utilization patterns.

4.2.2 Trajectory of PAC Service Use in Participating Initial Collection Markets

Although episodes of PAC are not the specific focus of the work under the PAC-PRD, past analyses by RTI and ASPE looking at trajectories of utilization within PAC episodes are useful for understanding whether the data collected in the PAC-PRD reflect the range of utilization patterns nationally. **Tables 4-5a** and **4-5b** show the top patterns of PAC utilization for beneficiaries in the PAC-PRD initial phase of data collection. Table 4-5a shows the trajectory for beneficiaries who initiated care in an acute hospital, and Table 4-5b shows the trajectory for beneficiaries who initiated care in HHAs, along with the rank of each pattern type compared with national data. These findings indicate that the data collected in the PAC-PRD are reflective of the types of beneficiaries and patterns of use nationally.

The most common episode trajectories observed for beneficiaries in the PAC-PRD data collection were similar to those in national data, although the PAC-PRD data reflected higher use of IRFs and LTCHs because of the intentional oversampling of these types of providers. The PAC-PRD data included beneficiaries initiating care in HHAs and in acute hospitals. The home health users who were not hospital discharges may have had different levels of acuity than those using HHAs after an acute hospitalization and are an important population to capture in these analyses. The PAC-PRD data also included a small number of beneficiaries initiating care in IRFs and LTCHs. Although these beneficiaries were a relatively small proportion of PAC users,

their use has been observed in analysis of national data, and it is important to note that these types of beneficiaries were represented in the PAC-PRD data collection.

In addition to the more common use trajectories, the PAC-PRD data also included beneficiaries with longer episodes of care and four or more settings of care. These beneficiaries may have had multiple acute hospitalizations and transitioned through several levels of care. Although the design of the PAC-PRD did not follow patients over time, but instead captured characteristics of patients at one point in their episode, each CARE assessment may reflect a different point in a patient's episode. For example, of the beneficiaries with the most common episode pattern in the PAC-PRD sample, acute-SNF-HHA, 7 percent of beneficiaries had CARE data collected in the acute setting, 59 percent in the SNF setting, and 33 percent in the HHA setting. Similarly, for beneficiaries with long trajectories of care, CARE data were collected at the beginning of the care trajectory for some and at the end of the trajectory for others. This result indicates that the CARE data used in the analyses included information on a range of types of patients and that data were collected at different points along the trajectory of service for different patients' use, ensuring that the data represent a range of patient complexity and utilization patterns. These data will provide important information on case mix across an episode of care.

4.3 Provider Selection and Recruitment

Table 4-6 provides a summary of the sample and the number of providers of each type that agreed to participate in the first phase of the study by market area. A total of 140 providers participated in the data collection across all initial data collection market areas. For the second phase we recruited an additional 66 providers to participate in the data collection. Within each market area, we targeted specific numbers of each type of provider according to the characteristics of the market. For example, we targeted LTCHs and freestanding IRFs in the "high-PAC-supply" areas, including Dallas, Louisville, and Boston, and we targeted greater numbers of SNFs and HHAs in the "low-PAC-supply" areas, such as Wilmington, Columbia, and Rochester. Acute hospitals, SNFs, and HHAs were targeted in all market areas, although we specifically targeted SNFs treating high-intensity patients in low-PAC areas. LTCHs and IRFs were oversampled in the provider selection to provide sufficient sample size of different types of cases treated in these settings. In the initial phase of data collection, one acute care provider was targeted for recruitment within each market, with the primary analytic focus of examining PAC discharge placement. In the supplemental phases of data collection, additional acute care providers were recruited, with a focus on capturing additional information on the treatment of the highly medically complex in step-down and other units within acute care providers. Some acute providers in the supplemental phase participated with more than one type of unit within their facility.

Providers were enrolled in the demonstration through RTI's recruitment efforts, which included mailings to providers of each targeted type and follow-up calls with chief executive officers, leadership teams, and nursing staff explaining the policy relevance of this work and the details of the data collection process. Participation in the demonstration was voluntary and included a \$3,000 stipend to defray some of the costs of data collection. The study did not include incentives to participate that would alter patterns of care.

Other factors considered in the provider sampling were the distribution of freestanding compared with hospital-based providers. Historically, hospital-based PAC providers have received the “sicker” group of patients. These patients can be transferred with relative ease, allowing hospitals to discharge a less stable patient, but with physician continuity. In addition, chain membership was considered during recruitment to ensure that the sample included both independent providers and representatives from some of the Nation’s major chains, because they account for large shares of patient treatments. The assessment data (Minimum Data Set [MDS], Inpatient Rehabilitation Facility Patient Assessment Instrument [IRF-PAI], and Outcome and Assessment Information Set [OASIS]) and case-mix data from claims (e.g., resource utilization groups [RUGs] for SNFs) were also used to ensure that providers targeted for recruitment would provide sufficient volume of cases and variety in types of cases, both in diagnoses and complexity. For example, we wanted to ensure that the sample included SNFs and HHAs treating medically complex cases, SNFs treating rehabilitation cases, and IRFs treating less intensive rehabilitation cases and ventilator weaning cases. Within providers that agreed to participate, we targeted a variety of units to ensure that our sample included patients with a range of diagnoses, both medical and surgical, such as stroke, rehabilitative diagnoses, pneumonia, and other respiratory diagnoses. A range of patient populations is represented in the units enrolled in the study. Study units across participating providers included stroke and neurology, cardiac care, orthopedic and rehabilitation, pulmonary and ventilator, and brain injury, as well as general medical-surgical units.

4.4 Data Collection

Two types of data were collected in the participating providers. First, all providers, including both acute hospitals and PAC providers, collected the standardized CARE assessment items. This process provided standardized measures of severity for each of the enrolled beneficiaries. Second, PAC providers collected CRU data, which provided staff-time measures for treating each of the enrolled beneficiaries. Site coordinators were identified for each participating site to manage the day-to-day logistics of data collection and data entry. Monthly coordinator calls provided an opportunity for site coordinators to communicate with each other and with RTI and to receive clarification on assessment items or data collection processes.

4.4.1 CARE Data Collection

The CARE data provided a standard way to measure patient medical, functional, cognitive, and social support factors. CARE data were collected by acute providers at the point of discharge and by HHAs, SNFs, and LTCHs upon admission and on discharge for a 9-month window in the initial phase of data collection. CARE data were collected at admission and discharge for all PAC providers or at admission and discharge from the relevant unit types for acute providers for a 6-month window in the supplemental phase of data collection. Data collection was initiated across market areas on a staggered basis beginning in March 2008 and continuing through 2010.

A Web-based application was created for electronic receipt of CARE data submitted by participating providers. The CARE tool was designed to be collected either on paper or directly through an Internet-based application. CMS’ vision in developing CARE was to move from multiple incompatible assessment instruments to one standardized set of clinically relevant data. Similarly, in developing a system that allows for the electronic transfer of patient assessment

information, CMS' vision was of fostering standardization and allowing for more consistent and informative communication between CMS and providers, as well as within providers, for the purposes of improving care transitions. The CARE application applies federally and nationally recognized health information technology standards. Use of broadly adopted health IT standards will allow for the safe, secure electronic exchange of critical health information among authorized users.

Before the start of data collection, RTI worked closely with a lead site coordinator and a backup site coordinator at each site. The site coordinators were responsible for overseeing the completeness, accuracy, and timing of the data collected. RTI worked with these staff members to incorporate the CARE data collection into their workflow and helped identify the appropriate staff to complete the items. Data collection periods began with a 1-day in-person intensive training of all coordinators (initial and supplemental) in a local market. The clinicians were trained in how to use the CARE items properly; access the Web-based CARE application; monitor the quality of the data they were collecting; and access the project resources available to them, including a Web-based coordinators' site, monthly coordinators' meetings, and the project help desk staffed by the Rehabilitation Institute of Chicago. The training was designed to draw comparisons with their current workflow practices, including the assessment items already used to admit patients to their care.

The 1-day training sessions were followed by site visits to each of the 140 initial data collection phase providers and the 66 supplemental data collection phase providers by a team of clinical and interview staff. Clinical team members conducted in-service trainings with the staff working on the participating units. Management teams at each site were interviewed about the populations they treat and their current methods for measuring case mix, planning staffing, and monitoring quality. This process also gave the organization's leadership an opportunity to ask questions and comment on the effort.

Data collection models varied across the providers in their approach for conducting the assessments. Each organization chose the data collection model that best reflected its individual work practices. The varying approaches and different types of staff used to complete the CARE assessments were consistent with CMS policy of allowing individual providers to identify the appropriate person in their setting to complete a standardized assessment. Some organizations, such as HHAs, used one assessor for each patient. Other organizations, such as IRFs, used different staff to complete different sections of the tool. All assessors collecting data in this project were licensed professionals. Nurses almost always completed the medical items, but the impairments and functional items were completed by nurses; physical therapists; occupational therapists; and, when appropriate, speech pathologists. The cognitive section was completed by nurses, occupational therapists, and case managers, depending on the individual facility. A total of 53,952 finalized CARE assessments were collected through December 31, 2010. Of these, 25,412 were admission assessments; 26,128 were discharges; 1,732 were interim assessments; and 680 were expired assessments.

Tables 4-7a and **4-7b** show the distribution by provider type of the finalized CARE assessments in the sample, by type of assessment, by phase of data collection. There are, in total, across all assessment types, 14,602 assessments from IRF providers; 11,374 from HHAs; 11,982 from SNFs; 8,560 from LTCHs; and 2,789 from acute providers, plus 4,645 assessments from

high-acuity patients in acute care hospitals (including step-down units). The specific sample used in each analysis is discussed in the relevant findings chapters. All samples excluded assessments for patients enrolled in a Medicare health maintenance organization.

4.4.2 CRU Data Collection

The CRU data collection effort was designed to address the mandate within this demonstration to measure patient-specific resource needs across PAC settings. It is not possible to measure patient-specific variable costs associated with different patients using administrative data in PAC settings. Therefore, to address this issue, the primary focus of CRU was staff-time measurement, capturing variations in types of staff, licensure levels, and total time spent with individual patients. To collect these data, RTI used data collection instruments designed to be completed by each staff person engaged in direct patient care in the participating providers. In the supplemental phase of data collection, the CRU data collection was expanded to also examine patient-specific staff time in selected medical care units and step-down units within acute providers. CRU data from the supplemental phase of data collection were not analyzed for this report, but they will be analyzed as part of other CMS work.

An important goal of this demonstration was to determine patient characteristics that drive differences in fixed and variable costs of PAC across settings. An individual's costs will vary by the patient acuity (both medical and functional, including cognitive) and the fixed costs associated with the provider type (setting) needed to deliver the appropriate treatment resources. The patients' variable costs (or resource use) were derived from the CARE tool data collection and the CRU data collection, as well as from additional charge information from the claims data. Their fixed costs were based on the provider-level costs associated with treatment at a particular level of care.

All staff on participating units were asked to track their time with patients. Each staff person who was engaged in direct patient care on each day during the data collection period used a pencil-and-paper data collection instrument (CRU data collection tool) to report time spent with the patient or on behalf of the patient. Total staff time included all direct care staff time and support staff time directly involved in the care of specific patients. Therapy staff time was reported in individual sessions and in sessions with two or more patients (e.g., groups or concurrent sessions). These minutes were allocated to the relevant patients. Time staff spent with patients on the participating units but not with any specific patient (e.g., team meeting) was allocated to individual patients according to that patient's share of individual time spent with that staff person. Non-physician staff, such as a dietician or social worker who treated study patients in the unit, were asked to sign a consultant log identifying their discipline and the time spent with individual patients. Staff time was then summed for each individual patient across all staff forms by occupational category to create a total staff time per patient-day.

In contrast, for HHAs, the resource use data were collected from both home health claims and home health staff treating a PAC-PRD patient. The information available from home health claims is a good estimate of direct patient care costs, although time estimates from HHA claims do not reflect non-face-to-face, patient-specific time. For HHAs, CRU data were obtained primarily from claims using the visit counts and minutes associated with each type of HHA service (therapy, nursing, aides, and social workers). Home health providers also collected

patient-specific staff time data using CRU, with the goal of better understanding non-face-to-face, patient-specific time, including travel time and documentation time.

In addition to the unit staff time and consultant time, imaging and other diagnostic tests or complex treatments were tracked in an ancillary log kept on the unit. The treatments and tests included on the ancillary log were selected because they are generally high cost and it was important to capture the resource use associated with these tests and treatments. Providers were trained to write in other tests or treatments not listed on the form if they felt that these tests and treatments were resource intensive. The major treatments section of the CARE tool also collected information on whether a treatment was used during the stay. Medications were included in the CARE tool as well, but this section was optional for the purposes of the demonstration.

Although participating providers collected CARE assessment data on 20–25 patients a month in the data collection period, CRU data collection was limited to three 2-week periods in the initial phase of data collection and to two 2-week periods in the supplemental phase of data collection. These CRU periods included weekends. Each round of CRU data collection was separated by approximately 2 months. Using this approach, we were able to collect information representing patient experiences at all points in the care trajectory, because in these three windows of time we captured patients at the beginning, middle, and end of their stays and at various points of their PAC episodes. Data collection began after in-person training was conducted. The first CRU data collection began in August 2008.

In-person trainings were conducted at each provider before the CRU data collection to train staff on the data collection tool and to meet with coordinators to provide additional support. RTI also provided additional webinar or teleconference trainings when necessary. Inpatient providers collected data on all patients on the study units (both Medicare and non-Medicare patients) to simplify data collection and eliminate the need for staff to identify Medicare patients. Site coordinators also provided a report of the daily total census and daily Medicare census on the unit for each of the CRU data collection days. HHAs collected CRU data only on study patients.

The CRU data sets used for analyses in this report include all data submitted as of April 2010. In total, 107 providers submitted CRU data. This number is less than the 140 providers who submitted CARE assessment data because of resource constraints or organizational issues that developed at some providers during the data collection period. By design, CRU was not collected in acute care hospitals in the initial phase of data collection. The CRU data collected during the supplemental phase of data collection were not included in this set of analyses but will be analyzed under another CMS project. A total of 15 LTCHs, 26 IRFs, 35 SNFs, and 31 HHAs submitted CRU data. Some providers collected fewer than three rounds of CRU data in the initial data collection phase. Therefore, more data were available on some providers than on others. HHA data were also supplemented with visit data in the HHA claims. Therefore, for HHAs, our sample consists of every day in the first 60-day episode provided in a participating agency during which a face-to-face visit occurred. Using this method, 4,071 patients receiving home health were included in the HHA sample, with 58,123 total HHA days in their episodes of care.

Table 4-8 provides estimates of the number of patient admissions, patient-days, and days per patient stratified by setting in our CRU data sample. The CRU sample includes 6,705 total admissions from all settings and almost 21,600 patient-days in all settings except HHAs. Stratified by setting (excluding HHAs), IRFs and SNFs had the most number of patient-days contributing to our sample size (8,256 and 6,691, respectively). LTCHs had almost as many patient-days, with 6,645 days.

Figure 4-1
Nationwide map of initial collection markets and supplemental collection cities



Table 4-1
Count of providers, by provider type, by initial collection market, 2006

Market	Region	Urban status	PAC resource	Acute	LTCH	Freestanding IRF	Hospital unit IRF	Freestanding SNF	Hospital unit SNF	Freestanding HHA	Hospital-based HHA
Boston	East	Urban	High	103	19	12	13	626	22	175	18
Chicago	Midwest	Urban/ suburban	High	140	14	6	54	518	26	455	43
Columbia	Midwest	Rural	Low	29	1	2	9	130	7	27	13
Dallas	South	Urban	High	101	19	10	25	325	6	555	17
Lakeland/ Tampa	Southeast	Suburban	Low	58	5	2	12	285	6	207	13
Lincoln ¹	Central	Rural	High	52	2	1	4	122	5	34	30
Louisville	Central	Suburban	High	72	10	6	10	286	23	62	31
Rochester	East	Suburban	Low	44	0	0	13	137	23	31	9
San Francisco ²	West	Urban	High	95	2	0	22	322	39	81	23
Seattle ³	West	Urban	Low	40	2	0	14	143	3	18	12
Sioux Falls	Central	Rural	Low	43	1	0	3	90	13	18	24
Wilmington	Southeast	Rural	Low	21	0	0	4	58	9	24	5

¹ Although the market included both Lincoln and Omaha, this table shows only the number of providers in the 2-hour driving distance of Lincoln, Nebraska.

² Although the market included both San Francisco and Sacramento, this table shows only the number of providers in the 2-hour driving distance of San Francisco, California.

³ Although the market included both Seattle and Portland, this table shows only the number of providers in the 2-hour driving distance of Seattle, Washington.

NOTE: HHA = home health agency; IRF = inpatient rehabilitation facility; LTCH = long-term care hospital; PAC = post-acute care; SNF = skilled nursing facility.

SOURCE: RTI analysis of 2006 Provider of Service data.

Table 4-2
Percentage of beneficiaries discharged to post-acute care and first site of care, by market area, 2006

Market	Number of beneficiaries with index acute hospitalizations	Percentage of beneficiaries discharged to PAC	Percentage discharged to LTCH ¹	Percentage discharged to IRF ¹	Percentage discharged to SNF ¹	Percentage discharged to HHA ¹	Percentage discharged to HOPD ¹
National 5% sample ²	310,628	35.2	2.0	10.3	41.1	37.4	9.1
Boston	184,578	47.8	3.2	5.9	47.0	38.3	5.6
Chicago	282,584	36.9	1.4	11.7	44.8	34.1	8.0
Columbia	37,695	35.3	0.3	7.9	43.8	37.7	10.3
Dallas	119,148	29.6	8.0	20.8	15.7	45.9	9.7
Lakeland/Tampa	165,498	39.5	0.6	5.7	44.0	43.4	6.3
Lincoln	37,737	35.7	3.6	4.4	42.5	27.0	22.5
Louisville	106,223	35.0	1.5	11.4	45.3	30.3	11.5
Omaha	40,076	34.8	3.5	6.2	40.5	27.9	22.0
Portland	29,674	31.2	0.0	4.3	51.7	34.7	9.3
Rochester	61,007	44.1	0.0	4.6	46.7	43.1	5.6
Sacramento	110,221	28.6	1.2	4.9	28.1	56.5	9.3
San Francisco	116,753	28.7	1.1	5.3	30.6	53.4	9.5
Seattle	60,949	35.0	0.3	6.7	51.6	29.7	11.7
Sioux Falls	14,912	30.9	0.4	6.3	49.1	21.8	22.5
Wilmington	42,713	30.2	0.3	9.7	38.3	43.9	7.7

¹ First site of PAC.

² Note that national estimates were based on the Medicare 5 percent sample. Market-level estimates were based on 100 percent of acute initiated episodes in the market area.

NOTE: Based on episode analysis conducted for the Office of the Assistant Secretary for Planning and Evaluation using 2006 Medicare claims data. Episodes were defined as starting with an index hospitalization and ending with the last hospital or PAC claim before a 60-day gap in acute and PAC service use. Independent therapists billing separately under Part B are not included in the episode definition used for this table. HHA = home health agency; HOPD = hospital outpatient department; IRF = inpatient rehabilitation facility; LTCH = long-term care hospital; PAC = post-acute care; SNF = skilled nursing facility.

SOURCE: RTI analysis of 2006 Medicare claims (bldm012).

**Table 4-3
Market area episode descriptives, 2006**

Market	Mean index acute LOS	Mean index acute payment	Mean PAC LOS	Mean PAC payment	Mean episode LOS	Mean episode payment
National 5% sample	6.8	\$10,297	74.7	\$16,058	82.6	\$27,886
Boston	6.1	\$9,866	71.7	\$16,274	78.5	\$26,140
Chicago	6.5	\$10,144	68.0	\$16,523	75.4	\$26,667
Columbia	6.1	\$8,388	58.5	\$11,728	65.6	\$20,115
Dallas	6.4	\$9,790	98.3	\$18,405	106.2	\$28,195
Lakeland/Tampa	6.3	\$8,702	60.3	\$13,239	67.4	\$21,940
Lincoln	6.0	\$9,396	57.3	\$11,638	64.4	\$21,034
Louisville	6.4	\$9,064	64.3	\$14,078	71.7	\$23,142
Omaha	6.0	\$9,497	56.9	\$11,568	64.2	\$21,065
Portland	6.1	\$11,007	54.9	\$11,793	62.4	\$22,799
Rochester	6.9	\$9,255	61.4	\$10,991	69.1	\$20,245
Sacramento	6.4	\$13,038	50.5	\$13,349	58.9	\$26,387
San Francisco	6.5	\$14,078	51.6	\$14,446	60.1	\$28,524
Seattle	5.7	\$11,082	55.7	\$13,672	62.6	\$24,754
Sioux Falls	5.2	\$7,161	48.9	\$7,662	55.7	\$14,823
Wilmington	7.1	\$8,559	65.7	\$12,283	73.9	\$20,842

NOTE: Based on episode analysis conducted for the Office of the Assistant Secretary for Planning and Evaluation using 2006 Medicare claims data. Episodes were defined as starting with an index hospitalization and ending with the last hospital or PAC claim before a 60-day gap in acute and PAC service use. Independent therapists billing separately under Part B are not included in the episode definition used for this table. Standardized payments are reported here to remove the effects of payment adjustments caused by geography or other policy considerations. LOS = length of stay; PAC = post-acute care.

SOURCE: RTI analysis of 2006 Medicare claims (bldm014).

Table 4-4a
Market area episode descriptives, by service type (HHA, IRF, and SNF), 2006

Market	% with at least one HHA claim	Mean HHA visits	Mean HHA payment	% with at least one IRF claim	Mean IRF LOS	Mean IRF payment	% with at least one SNF claim	Mean SNF LOS	Mean SNF payment
National 5% sample	60.3	25.9	\$3,916	11.7	13.9	\$16,289	47.8	37.3	\$11,242
Boston	69.7	23.4	\$3,551	6.9	16.6	\$18,781	54.2	36.3	\$12,186
Chicago	57.2	20.7	\$3,579	13.2	13.4	\$16,512	51.9	38.4	\$12,138
Columbia	57.1	18.6	\$2,857	9.0	14.7	\$16,215	50.4	36.0	\$9,159
Dallas	68.6	48.0	\$6,171	23.2	13.3	\$14,916	20.1	40.6	\$11,333
Lakeland/Tampa	69.8	19.8	\$3,264	6.4	14.3	\$15,276	49.6	37.5	\$12,141
Lincoln	43.1	18.2	\$2,830	5.0	14.2	\$15,997	50.4	34.4	\$9,441
Louisville	53.2	21.8	\$3,276	13.1	13.6	\$15,026	52.1	38.7	\$10,736
Omaha	43.3	18.7	\$2,848	7.0	12.9	\$14,549	48.4	34.0	\$9,345
Portland	58.2	14.6	\$2,880	4.8	12.3	\$17,543	55.4	30.6	\$10,535
Rochester	65.4	20.3	\$2,813	5.0	15.5	\$17,509	52.0	32.7	\$8,801
Sacramento	71.3	14.3	\$3,450	5.4	14.3	\$23,837	31.4	35.9	\$14,533
San Francisco	70.1	14.6	\$3,646	5.9	14.4	\$25,571	34.0	34.4	\$14,738
Seattle	53.1	15.0	\$3,086	7.3	10.8	\$17,275	57.3	34.0	\$11,580
Sioux Falls	32.4	17.5	\$2,529	7.3	10.3	\$11,735	54.2	24.3	\$6,321
Wilmington	62.3	21.0	\$3,421	10.8	12.6	\$15,818	43.5	38.8	\$10,324

NOTE: HHA visits are calculated from the start of services to the end of billed services. Note that this period may cover multiple HHA episodes. Episodes were defined as starting with an index hospitalization and ending with the last hospital or PAC claim before a 60-day gap in acute and post-acute care (PAC) service use. Independent therapists billing separately under Part B are not included in the episode definition used for this table. Mean payment is based on patients initiating episodes with an index acute hospitalization. This analysis does not include beneficiaries entering PAC services without an index acute hospitalization. Standardized payments are reported here to remove the effects of payment adjustments caused by geography or other policy considerations. HHA = home health agency; IRF = inpatient rehabilitation facility; LOS = length of stay; SNF = skilled nursing facility.

SOURCE: RTI analysis of 2006 Medicare claims (bldm014).

Table 4-4b
Market area episode descriptives, by service type (LTCH, HOPD, and readmission), 2006

Market	% with at least one LTCH claim	Mean LTCH LOS	Mean LTCH payment	% with at least one HOPD claim	Mean HOPD units	Mean HOPD payment	% with at least one acute hospital readmission claim	Mean acute hospital readmission LOS	Mean acute hospital readmission payment
National 5% sample	2.9	32.5	\$38,559	22.9	45.3	\$1,258	30.5	11.5	\$15,636
Boston	4.3	30.8	\$29,969	17.4	33.3	\$1,119	27.1	10.8	\$16,327
Chicago	2.1	33.2	\$44,105	21.9	43.7	\$1,286	29.2	11.2	\$16,364
Columbia	0.5	38.9	\$51,901	24.0	43.3	\$1,105	26.7	10.6	\$13,055
Dallas	10.6	30.6	\$34,687	24.0	43.7	\$1,366	28.9	11.4	\$15,338
Lakeland/Tampa	0.9	32.8	\$39,312	19.0	49.2	\$1,344	25.6	10.8	\$13,043
Lincoln	4.6	25.3	\$30,962	38.6	36.1	\$928	23.4	9.7	\$13,164
Louisville	2.2	32.4	\$41,684	24.8	38.9	\$1,077	26.6	10.5	\$13,535
Omaha	4.4	25.2	\$31,196	38.2	35.8	\$917	23.6	9.7	\$13,072
Portland	0.0	44.5	\$83,101	21.0	30.4	\$1,024	22.2	8.6	\$14,428
Rochester	0.0	45.5	\$61,443	14.4	23.3	\$814	25.4	12.4	\$14,046
Sacramento	1.5	30.8	\$41,016	18.7	38.8	\$1,085	22.5	10.3	\$18,830
San Francisco	1.4	31.7	\$42,487	19.4	39.7	\$1,212	22.9	10.6	\$19,874
Seattle	0.5	35.8	\$67,080	24.4	31.5	\$1,092	23.0	8.6	\$15,296
Sioux Falls	0.5	24.1	\$36,619	37.3	41.0	\$691	21.3	7.8	\$9,975
Wilmington	0.5	37.9	\$37,059	19.1	56.4	\$1,342	26.5	12.1	\$13,233

NOTE: HOPD units as reported on the outpatient department claim. Episodes were defined as starting with an index hospitalization and ending with the last hospital or post-acute care (PAC) claim before a 60-day gap in acute and PAC service use. Independent therapists billing separately under Part B are not included in the episode definition used for this table. Mean payment is based on patients initiating episodes with an index acute hospitalization. This analysis does not include beneficiaries entering PAC services without an index acute hospitalization. Standardized payments are reported here to remove the effects of payment adjustments caused by geography or other policy considerations. HOPD = hospital outpatient department; LOS = length of stay; LTCH = long-term care hospital.

SOURCE: RTI analysis of 2006 Medicare claims (bldm014).

Table 4-5a
Top 10 episode patterns in the initial PAC-PRD sample
for beneficiaries initiating service use in an acute hospital

Episode pattern ¹	Rank in PAC-PRD sample	n	Percentage of acute-initiated episodes in PAC-PRD sample	Rank in national data ²	Percentage of acute-initiated episodes in national data
ASH	1	1,282	9.0	3	8.6
AH	2	1,198	8.4	1	22.9
AIH	3	950	6.7	8	2.7
AS	4	636	4.5	2	13.9
A	5	488	3.4	—	—
AIO	6	302	2.1	16	0.9
AHA	7	282	2.0	5	3.7
AI	8	280	2.0	18	0.7
AL	9	236	1.7	26	0.5
ALH	10	201	1.4	41	0.2

¹ The sample for this analysis was limited to beneficiaries with Continuity Assessment Record and Evaluation (CARE) assessment data that matched to Medicare claims data and for whom an initiating event was identified. Episode pattern is based on a 30-day variable-length episode definition after an acute hospital claim that followed a 30-day period without acute, IRF, LTCH, SNF, or HHA service use. The last claim in an episode is the last claim before a 30-day gap in acute, IRF, LTCH, SNF, HHA, or therapy service use. Each letter indicates use of a type of service, but note that a single letter may represent one claim or multiple claims for services of the same type: A = acute hospital; H = home health agency; I = inpatient rehabilitation facility; L = long-term care hospital; O = outpatient department therapy; S = skilled nursing facility; T = independent therapist.

² Note that these analyses focused only on beneficiaries using PAC services and therefore did not include beneficiaries with an acute hospitalization only. The episode patterns in the top 10 nationally not shown in this table are Rank 4 = AO; Rank 6 = AT; Rank 7 = ASO; Rank 9 = ASAS; Rank 10 = AHO.

NOTE: PAC-PRD = post-acute care payment reform demonstration; SNF = skilled nursing facility.

SOURCE: RTI analysis of Medicare claims. National data presented in Gage, B., Morley, M., Ingber, M., and Smith, L.: Post Acute Care Episodes Expanded Analytic File. Final report prepared for Assistant Secretary for Planning and Evaluation. Waltham, MA: RTI International, April 2011.

Table 4-5b
Top five episode patterns in the initial PAC-PRD sample for
beneficiaries initiating service use in an HHA

Episode pattern ¹	Rank in PAC-PRD sample	n	Percentage of HHA-initiated episodes in PAC-PRD sample	Rank in national data	Percentage of HHA-initiated episodes in national data
H	1	816	37.4	1	70.1
HA	2	90	4.1	2	5.9
HASH	3	78	3.6	7	1.3
HAH	4	72	3.3	3	3.7
HAS	5	66	3.0	5	1.8

¹ The sample for this analysis was limited to beneficiaries with Continuity Assessment Record and Evaluation (CARE) assessment data that matched to Medicare claims data and for whom an initiating event was identified. Episode pattern is based on a 30-day variable-length episode definition after an HHA claim that followed a 30-day period without acute, IRF, LTCH, SNF, or HHA service use. The last claim in an episode is the last claim before a 30-day gap in acute, IRF, LTCH, SNF, HHA, or therapy service use. Each letter indicates use of a type of service, but note that a single letter may represent one claim or multiple claims for services of the same type: A = acute hospital; H = home health agency; S = skilled nursing facility.

NOTE: PAC-PRD = post-acute care payment reform demonstration.

SOURCE: RTI analysis of Medicare claims. National data presented in Gage, B., Morley, M., Ingber, M., and Smith, L.: Post Acute Care Episodes Expanded Analytic File. Final report prepared for Assistant Secretary for Planning and Evaluation. Waltham, MA: RTI International, April 2011.

Table 4-6
Count of participating providers, by provider type, by market area

Market area	Acute	LTCH	IRF	SNF	HHA	Total
Initial high-PAC areas						
Boston	2	4	2	8	4	20
Chicago	1	2	5	6	5	19
Dallas	2	3	5	4	5	19
San Francisco/Sacramento	2	0	2	6	2	12
Seattle/Portland	0	1	1	3	2	7
Lincoln/Omaha/Sioux Falls	2	1	2	4	4	13
Louisville	1	2	4	4	2	13
Subtotal, high-PAC areas	10	13	21	35	24	103
Initial low-PAC areas						
Columbia	1	0	0	2	5	8
Lakeland/Tampa	1	2	2	1	3	9
Rochester	1	0	1	3	6	11
Wilmington	2	0	2	2	3	9
Subtotal, low-PAC areas	5	2	5	8	17	37
Total initial data collection phase	15	15	26	43	41	140
Supplemental areas						
New Hampshire/Maine	2	0	1	1	0	4
New York/New Jersey/ Pennsylvania	2	1	5	1	1	10
Virginia	3	2	3	4	0	12
Maryland/D.C.	3	5	2	7	2	19
North Carolina	5	0	0	3	0	8
Ohio/Michigan	3	3	1	1	0	8
California	2	2	1	0	0	5
Total supplemental data collection phase	20	13	13	17	3	66
Total initial and supplemental data collection phases	35	28	39	60	44	206

NOTE: The total number of participating providers was 140. HHA = home health agency; IRF = inpatient rehabilitation facility; LTCH = long-term care hospital; PAC = post-acute care; SNF = skilled nursing facility.

SOURCE: RTI PAC-PRD data collection.

Table 4-7a
Initial sample CARE assessment counts by assessment type, by provider type

Overall	HHA	SNF	IRF	LTCH	Acute	Total
Admission	5,292	4,850	5,831	3,400	—	19,373
Discharge	4,587	4,305	5,601	2,944	2,789	20,226
Expired	33	146	11	300	—	490
Interim	752	296	52	371	—	1,471
Total	10,664	9,597	11,495	7,015	2,789	41,560

NOTE: Dash (—) indicates that these assessment types were not collected at acute providers. CARE = Continuity Assessment Record and Evaluation; HHA = home health agency; IRF = inpatient rehabilitation facility; LTCH = long-term care hospital; SNF = skilled nursing facility.

SOURCE: Microstrategy report 2, December 28, 2010.

Table 4-7b
Supplemental providers CARE assessment counts by assessment type, by provider type

Overall	HHA	SNF	IRF	LTCH	Acute	Total
Admission	332	1,204	1,549	775	2,179	6,039
Discharge	318	1,040	1,543	626	2,375	5,902
Expired	1	39	3	73	74	190
Interim	59	102	12	71	17	261
Total	710	2,385	3,107	1,545	4,645	12,392

NOTE: CARE = Continuity Assessment Record and Evaluation; HHA = home health agency; IRF = inpatient rehabilitation facility; LTCH = long-term care hospital; SNF = skilled nursing facility

SOURCE: Microstrategy report 2, December 28, 2010.

Table 4-8
Number of CRU patients, patient-days, and days per patient,
by setting

Setting	Admissions	Patient-days	Mean days per patient
All settings	6,705	79,715	11.89
LTCH	728	6,645	9.13
IRF	1,106	8,256	7.46
SNF	800	6,691	8.36
HHA	4,071	58,123	14.28

NOTE: Days per patient for HHAs were based on claims, not CRU data collection as for the other settings. CRU = cost and resource utilization; HHA = home health agency; IRF = inpatient rehabilitation facility; LTCH = long-term care hospital; SNF = skilled nursing facility.

SOURCE: RTI CRU data, April 30, 2010.