February 2014

Developing Outpatient Therapy Payment Alternatives: Final Report

Task 5, Deliverable 2

Prepared for

Ann Meadow, ScD
Centers for Medicare & Medicaid Services
Center for Medicare & Medicaid Innovation
Research and Rapid Cycle Evaluation Group
Mail Stop WB-06-05
7500 Security Boulevard
Baltimore, MD 21244-1850

Prepared by

Gregory Pope, MS
Peter Amico, PhD
Poonam Pardasaney, MS, DPT, ScD
Tracy Kline, PhD
Anne Deutsch, RN, CRRN, PhD
Walter Adamache, PhD
Melvin Ingber, PhD
Traci Gordon, BA
Jill A. Dever, PhD
RTI International

Alan Jette, PhD, Boston University
Barbara Gage, PhD, The Brookings Institution
Benjamin Silver, BA, Brown University

RTI Project Number: 0213520.000.005

[This page intentionally left blank.]

DEVELOPING OUTPATIENT THERAPY PAYMENT ALTERNATIVES: FINAL REPORT

by Jill A. Dever, RTI Project Director

Gregory Pope, Peter Amico, Poonam Pardasaney, Tracy Kline, Anne Deutsch,

Walter Adamache, Melvin Ingber, Traci Gordon,

Alan Jette (Boston University), Barbara Gage (The Brookings Institute), and

Benjamin Silver (Brown University)

Federal Project Officer: Ann Meadow, ScD

RTI International

CMS Contract No. HHSM-500-2012-00169G

February 2014

This project was funded by the Centers for Medicare & Medicaid Services under contract no. HHSM-500-2012-00169G.

RTI International is a trade name of Research Triangle Institute.

[This page intentionally left blank.]

CONTENTS

Appendices

List of Tables

1. INTRODUCTION

The Centers for Medicare & Medicaid Services (CMS) established a research project in 2008 titled Developing Outpatient Therapy Payment Alternatives (DOTPA) for the purposes of identifying, collecting, and analyzing therapy-related information tied to beneficiary need and resource use of outpatient therapy.[1] The research project spanned 6 calendar years and was conducted in two phases. The ultimate goal for the Phase I study (concluded January 2013) was to collect measures of functional status from Medicare beneficiaries. In Phase II (concluded February 2014), analyses of the data collected during Phase I were completed, and characteristics for alternatives to the current payment system were suggested.

This final report begins with some background information on current Medicare payment policy to provide context to the DOTPA project. We also highlight the relevant activities accomplished under DOTPA Phase I in ***Section 2***, including data collection and preliminary analyses. Next, we summarize the research completed under Phase II. The Phase II research is described in the next two sections—the analysis of the measurement properties of the data collection instrument is given in ***Section 3*** based on the *DOTPA Measurement Report* (Kline et al., 2014), and the analysis of the characteristics associated with resource utilization that may be used to form an alternative payment system is contained within ***Section 4*** based on the *DOTPA Payment Alternatives Report* (Amico et al., 2014b). ***Section 5*** contains caveats that should be considered when interpreting the DOTPA results along with lessons learned should CMS pursue further studies in this area in the future. We complete the report with suggestions for future research (***Section 6***) and an overall conclusion (***Section 7***).

---

[1] Throughout this report, the term “outpatient therapy” refers to services provided and paid under Medicare’s Part B therapy benefit.

[This page intentionally left blank.]

2. BACKGROUND

This section presents background on Medicare payment policy and on the Developing Outpatient Therapy Payment Alternatives (DOTPA) project, including a brief history of instrument development and primary data collection for the study.

2.1 Background on Medicare Therapy Benefit and Payment Policy

Physical medicine and rehabilitation is “a medical specialty concerned with the prevention, diagnosis, treatment, and management of disabling diseases, disorders, and injuries typically of a musculoskeletal, cardiovascular, neuromuscular, or neurological nature by physical means” (Merriam-Webster, Inc., 2012). In the Medicare program, the Social Security Act defines rehabilitation under one of three disciplines (Social Security Act of 1997):

- physical therapy (PT) – rehabilitative therapies related to physical function and impairment;
- occupational therapy (OT) – rehabilitative therapies focused on engagement in meaningful activities of daily living; and
- speech-language pathology (SLP) – rehabilitative therapies to address communication and swallowing disorders.

Though these disciplines appear to be clearly defined, their activities occasionally overlap. PT and OT, for example, both share an overall goal of improving and strengthening activities related to motion. PT typically focuses on the motions themselves, whereas OT focuses on applied activities of daily living.

Therapy services are provided at various stages of care including, but not limited to, inpatient, post-acute, outpatient, and long-term care. Medicare Part B rehabilitation services, usually provided on an outpatient basis, are the focus of the research discussed in this report. Part B therapy services were provided to 4,630,593 beneficiaries in 2010 (13.5 percent of the Medicare population), and comprised $5.6 billion in reimbursements in that year (3.5 percent of all Medicare benefit expenditures; Centers for Medicare & Medicaid Services [CMS], 2010). With an increase of 37.4 percent from 2006 (Silver and Dever, 2013), Part B therapy represents a small but rapidly growing percentage of overall Medicare expenditures, though the causes for this increase are not fully understood.

Part B therapy services are administered in various settings, addressing a wide range of beneficiary needs associated with, for example, medical complexity, living situations, and presence of caregivers. These settings include the following:

- hospital outpatient departments (HOPDs),
- rehabilitation agencies (sometimes referred to as outpatient rehabilitation facilities),

- comprehensive outpatient rehabilitation facilities,
- assisted living facilities (ALFs),[2]
- day rehabilitation (day rehab),[3]
- skilled/nursing facilities (S/NFs),[4]
- home health agencies,[5]
- therapists' private practices, and
- physicians' offices and offices of certain nonphysician practitioners.

Part B therapy services are typically sought for a variety of reasons such as (1) acute and chronic disease conditions; (2) injuries including those related to an acute hospital admission; or (3) in some cases, after a hospitalization once post-acute care benefits have been exhausted (CMS, 2012). Although these three needs are very different, they share a common theme: Medical issues for patients of Part B therapy are typically less severe than those of inpatients or most patients who are immediately post-acute.

Medicare beneficiaries receive outpatient rehabilitation under Part B, the portion of the Medicare benefit responsible for physician and outpatient hospital services, among other benefits (CMS, 2012). The reimbursement structure currently in place was signed into law in 1997 with the Balanced Budget Act. This legislation placed all Part B therapy services on the fee-for-service (FFS)–based Medicare Physician Fee Schedule and established a discipline-specific limit (therapy cap) on the amount of services a beneficiary could receive in a calendar year (Balanced Budget Act of 1997).[6] Before this act, physician offices and therapists in private practice were already reimbursed in this way, but several different mechanisms were employed across the other settings. The purpose of this shift was to standardize the method for which outpatient therapy services were reimbursed across *all* settings. Since 2006, a process for approving exceptions to the therapy caps has been periodically established or extended by Congress.

---

[2] ALFs are not recognized as Medicare providers; however, many do have outpatient rehabilitation services provided on site. This is achieved, among other ways, by a rehabilitation agency contracting with the facility to provide services on site, or by the facility being co-located with a nursing facility (i.e., a retirement community including a nursing facility) that provides therapy to outpatients.

[3] Providers classified as day rehab in this study are recognized as HOPDs by the Medicare program, and they self-identify as day rehab providers. Their services are furnished as outpatient hospital services, and they offer multiple therapeutic and medical services to beneficiaries during visits that may last several hours each.

[4] The therapy services furnished in private nursing facilities (NFs) are billed by the private therapist or other enrolled Medicare provider under contract. However, for the purposes of the analyses presented in the subsequent sections, we group skilled and other nursing facilities together under the acronym S/NF.

[5] Homebound Medicare beneficiaries may also receive therapy services under a home health plan of care paid under the home health prospective payment system. Therapy services paid under the home health payment system are outside the scope of analysis in this report.

[6] Until 2012, therapy services delivered in HOPDs were exempt from the caps.

2.2 Issues with Current Medicare Part B Therapy Payment Policy

Current Medicare policy sets the same annual therapy expenditure caps for all beneficiaries, consisting of one cap for PT and SLP combined ($1,920 in 2014) and a second, separate cap for OT ($1,920 in 2014). This can be viewed as inequitable because the caps are not adjusted for specific beneficiary needs for therapy services. Moreover, Medicare Part B therapy services are paid on an FFS basis. FFS payment may encourage the provision of more services, because more services increase provider revenue, and services up to the established cap are allowed without the exceptions process. Overutilization of services is a particular concern in therapy because clinical standards for the amount of therapy are often uncertain. The quantity of services provided may be sensitive to payment policy and provider profit incentives rather than to beneficiary need for services.

Case-mix adjustment of therapy payment can potentially help with both concerns.[7] First, this adjustment can enable a more equitable distribution of resources by establishing therapy caps that meet the therapy needs of the Part B eligible beneficiaries within the case-mix groups. Second, many alternatives to FFS payment (e.g., episode-based payment) involve bundling therapy services that will require case-mix adjustment of payment to address varying combinations of services used to treat eligible beneficiaries.[8]

2.3 The DOTPA Project

The DOTPA initiative began in 2008. At its completion, DOTPA spanned 7 calendar years. Development of the DOTPA instruments, provider and beneficiary recruitment, and data collection were conducted from January 2008 through January 2013 (Phase I) and are summarized as follows. Detailed information is publicly available in a series of reports cited in the project’s Year 5 report (Silver and Dever, 2013). Analyses of the DOTPA data along with Medicare claims, resulting in a series of recommendations, were conducted from September 2013 through January 2014 (Phase II) and are summarized in the subsequent sections of this final report.

The central goal of the DOTPA project is to collect data and conduct analyses that are helpful in defining the characteristics useful for case-mix adjustment of Medicare outpatient therapy payments. In particular, one element that clinicians believe to be critical in therapy case-mix measurement is patient functional status. Functional status is not routinely available from Medicare administrative data. Instead, functional status must be collected through clinician assessment and/or patient self-report. The DOTPA outpatient therapy case-mix measures may also be useful in establishing metrics for assessing post-therapy outcomes, although outcomes are not analyzed in this report.

---

[7] As defined by CMS, “the adjustment for the health condition, or clinical characteristics, and service needs of the beneficiary is referred to as the case-mix adjustment.” Within the research presented here, the suggestion would be to establish different payment allowances for beneficiaries grouped by relevant characteristics correlated with therapy needs and Part B resource usage.

[8] Payment for “bundled therapy services” refers to reimbursements given for a combination of therapy services used to treat a single illness or condition (http://innovation.cms.gov/initiatives/bundled-payments/). This differs from the FFS payment model, which reimburses the provider for each individual service used to treat a patient.

2.3.1 DOTPA CARE Assessment

Early work of the DOTPA project focused on the development of an instrument that could facilitate the efficient yet comprehensive collection of functional assessment information relevant to the needs of the Medicare Part B therapy population. The Continuity Assessment Record and Evaluation (CARE) instrument, originally designed to measure functional impairment in post-acute settings and tested in the CMS Post-Acute Care Payment Reform Demonstration (Gage et al., 2012), was selected as the template for the DOTPA instrument. Modifications were made to tailor the original CARE item set to the less clinically complex outpatient population studied under DOTPA. Two sections were added to generate a more complete picture of patients' needs: (1) a patient self-report section modified from materials developed for the Activity Measure for Post-Acute Care (AM-PAC[9]), and (2) communication and cognitive function items adapted from the National Outcomes Measurement System (NOMS[10]).

In addition to the three source instruments (CARE, AM-PAC, and NOMS), RTI and CMS involved practitioners, advocacy groups, and other stakeholders in the development of the DOTPA assessment tool through a technical expert panel meeting, several outreach activities, and an open-door forum. Organizations involved in one or more events included the following:

- American Physical Therapy Association,
- American Occupational Therapy Association,
- American Speech-Language-Hearing Association,
- American Medical Rehabilitation Providers Association,
- National Association for the Support of Long Term Care,
- American Health Care Association,
- TriAlliance of Health and Rehabilitation Professions,
- Uniform Data System for Medical Rehabilitation,
- Genesis Rehab Services, and
- Alexian Rehabilitation Hospital.

These efforts culminated in three DOTPA CARE assessment instruments intended to capture substantial differences in patient complexity for certain groups within the Part B therapy population:

1) a DOTPA CARE-C assessment form for patients living in the community,

---

[9] © CRE CARE, 2002–2009, http://www.crecare.com/am-pac/ampac.html.

[10] © American Speech-Language-Hearing Association, 1997–2013, http://www.asha.org/NOMS/.

2) a DOTPA CARE-F nursing home assessment form for patients who receive Part B therapy while residing in a nursing facility,[11] and

3) a DOTPA CARE-F day rehab assessment form for the more complex subset of the Part B outpatient therapy population who receive the comprehensive services of a day rehab[12] program.

For brevity, the instruments will be referred to as CARE-C and CARE-F, respectively, and as CARE assessment collectively in the remainder of the report. The two variants of CARE-F contain nearly all of the items on the CARE-C, as well as additional items tailored to capture information for possibly highly impaired patients who are typically not able to function in the community. The difference between the two variants of CARE-F includes sections in the nursing facility admission instrument on major medical treatments, such as central line management, hemodialysis (Section II, item C), and on persistent vegetative state (Section IV, item A), and a section in the day rehab admission instrument on social participation (Section II, item B.3).

2.3.2 Primary Data Collection Procedures

Primary data collection of patient assessments began in March 2011 and continued through June 2012 within a set of provider sites recruited for DOTPA. Provider recruitment began in late 2010 and continued on a rolling basis until early 2012. A staff member at each participant site was selected as site coordinator and acted as the primary point-of-contact with the DOTPA project team.

After the providers reviewed the details of the project and agreed to participate, the site coordinators received comprehensive Web-based training on the DOTPA data collection protocol and the relevant assessment instrument. Monthly group conference calls were also held with site coordinators to discuss beneficiary recruitment goals (established by the project team from historical patient counts and participant estimates of Medicare patient volume) and their progress toward these goals, to review data collection procedures to ensure high-quality assessment responses, and to address any questions or issues common to multiple providers. A project help-desk was established to assist the coordinators with any immediate issues and to answer questions; key help-desk discussions were summarized for the group at the monthly meetings.

Patient assessments were collected at most provider sites over a 6-month period. The number of assessments obtained varied across providers because of several factors including (1) differences in the number of patients treated during the data collection period across the participating sites; (2) the proportion of total patients using Medicare Part B; and (3) the availability of staff to collect and submit completed assessment forms. Some participating

---

[11] This patient population includes both long-term care residents and post-acute care patients who have exhausted their S/NF benefit under Part A.

[12] Day rehab is not an official provider licensure and as such is not clearly defined. Day rehab patients often have greater needs, and the services provided are typically more comprehensive and more frequent than standard outpatient care. Providers enrolled in the project as day rehab self-identified as such.

providers, for example, were smaller practices that did not have the resources to collect a large number of assessments. Practicing therapists, office staff, and the patients themselves all contributed to completion of the assessment. Most participating sites agreed to include all of their Part B patients, whereas some agreed to a systematic selection method that was less burdensome.[13]

The assessment was administered at both the beginning (admission assessment) and end (discharge assessment) of a course of therapy to capture the initial patient condition and a measure of change/improvement in the patient's status, respectively. Assessments were collected primarily on paper forms and then scanned into a database using TeleForm,[14] optical character recognition software. In later months of data collection, a Web-based system was deployed and several participating sites submitted assessments in this manner.

A total of 162 providers actively participated in DOTPA, with the largest number of participating sites coming from private practice (42.0 percent; ***Table 2-1***). Cumulatively, these providers submitted a total of 6,662 complete admission assessments that contained sufficient information for analyses—5,822 CARE-C assessments (***Table 2-2***) and 840 CARE-F assessments (***Table 2-3***)—linked to 6,490 unique beneficiaries (5,742 CARE-C and 748 CARE-F).[15] As noted previously, data from the admission assessments and not the discharge assessments were the primary focus of the DOTPA analyses summarized in this report.

Additional details on the design of the project and the data collection methodology are provided in an earlier report from the DOTPA project (Silver and Dever, 2013).

---

[13] Many of the participating sites were small practices that did not have extra time and overhead funding available to participate in the DOTPA data collection. The decision was made to make special accommodations for these sites because they represent an important share of the pool of outpatient therapy providers. Some sites were allowed to collect fewer assessments than their larger organizational counterparts; other providers were allowed to collect fewer assessments per month than the desired rate but continued data collection for an additional period of time to ensure sufficient overall numbers.

[14] Copyright © 2012 Autonomy, an HP Company.

[15] A total of 401 admission assessments out of 7,063 (5.7 percent; 338 CARE-C and 63 CARE-F) were deemed unusable because of, for example, incorrect Medicare Health Insurance Claim Numbers even after repeated attempts to obtain the corrected information.

Table 2-1
Total number of DOTPA-participating providers by setting

Setting	Providers participating in DOTPA Count	Providers participating in DOTPA Percent[1]
Total	162	100
Private practice	68	42
Hospital outpatient department	31	19
Outpatient rehabilitation facility	27	17
Skilled/nursing facility	21	13
Comprehensive outpatient rehabilitation facility	4	3
Assisted living facility	6	4
Day rehabilitation	5	3

NOTE:

* = Fewer than 11 cases.

[1] Setting-specific percentages may not sum to total because of rounding.

SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA) analytic files, 2014.

Table 2-2
Total number of DOTPA CARE-C admission assessments by setting and discipline

			Therapy discipline					
	Total		PT		OT		SLP	
Setting	Count	Percent[1]	Count	Percent[1]	Count	Percent[1]	Count	Percent[1]
Total	5,822	100	5,007	100	588	100	227	100
HOPD	1,450	25	1,090	22	210	36	150	66
ORF	998	17	904	18	80	14	14	6
CORF	42	1	*	*	*	*	34	15
ALF	255	4	157	3	76	13	22	10
S/NF	35	1	**	**	*	*	*	*
PP	3,042	52	2,819	56	**	**	*	*

NOTES:

* = Fewer than 11 cases.

** = Count suppressed to maintain confidentiality of cells with fewer than 11 cases.

ALF = assisted living facility; CARE-C = Continuity Assessment Record and Evaluation form for patients living in the community; CORF = comprehensive outpatient rehabilitation facility; HOPD = hospital outpatient department; ORF = outpatient rehabilitation facility; OT = occupational therapy; PP = private practice; PT = physical therapy; SLP = speech and language pathology; and S/NF = skilled/nursing facility.

[1] Setting-specific percentages may not sum to total because of rounding.

SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA) analytic files, 2014.

Table 2-3
Total number of DOTPA CARE-F admission assessments by setting and discipline

Setting	Count	Percent
Total	840	100
Skilled/nursing facility	655	78
Day rehabilitation facility	185	22

NOTE:

CARE-F = Continuity Assessment Record and Evaluation form for patients who receive Part B therapy while residing in a nursing facility or receiving day rehabilitation services.

SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA) analytic files, 2014.

2.3.3 Preliminary Medicare Claims Analysis

Beginning in the second year of the project, several preliminary claims analyses were conducted to gain a better understanding of Part B therapy utilization patterns and to help fulfill CMS's need for data to oversee the therapy benefit. These analyses covered therapy utilization from calendar years 2007–2010 and provided both helpful context and historical trends for future work on the project. Four reports were generated and are available from the Studies and Reports site on the CMS Therapy Services Web page at http://www.cms.gov (Kandilov, Lyda-McDonald, and Drozd, 2009; Lyda-McDonald, Drozd, and Gage, 2012; Silver et al., 2012; Silver, et al., 2013). In general, these analyses revealed a steady increase in utilization rates year after year and emphasized the difficulty of understanding beneficiary therapy needs with routine administrative data.

2.3.4 Construction of DOTPA Analytic Data Files

After completion of data collection, DOTPA assessment data were evaluated further for completeness,[16] quality, and a valid beneficiary Health Insurance Claim Number. The viable assessments were then assembled into four analysis data files in preparation for analysis—CARE-C admission, CARE-C discharge, CARE-F admission, and CARE-F discharge—corresponding to the DOTPA assessment instrument and the point when the assessment items were collected.

Derived variables created from combinations of items obtained from the CARE instruments were added to these data files. Two derived analytic files were also generated to capture the episode-level records for the CARE-C and CARE-F assessments (see ***Section 4.6*** of this report for a discussion of the episode-level analyses).

---

[16] As the DOTPA assessments were received, the data were checked manually for completeness and for errors that could be detected by visual inspection. If issues were identified, the provider site coordinator was asked to address the problems. An ideal data collection approach would have been an electronic data capture system with real-time response validation capabilities.

Data documentation and codebooks were generated for the six DOTPA data files (Amico et al., 2014a). These documents, along with the corresponding data files, were delivered to CMS at the end of the project, February 2014.

2.3.5 Preliminary Analysis: DOTPA CARE Measurement Properties

Analyses were conducted with the analytic files to assess the measurement properties of the DOTPA assessment items using descriptive and Rasch techniques. A set of scales to describe several dimensions of functional ability was constructed from the data. The analyses and findings are detailed in the *DOTPA Measurement Report* (Kline et al., 2014) and are summarized in ***Section 3*** of this report.

2.3.6 Analysis for Characteristics for an Alternative Payment System

Upon completion of the measurement analyses, the newly formed scales and the DOTPA CARE assessment information were combined with Medicare claims data. The augmented analytic files were used to investigate therapy utilization patterns and to identify possible characteristics of alternatives to the current Medicare Part B reimbursement structure.

We also analyzed 100 percent of Medicare claims for Part B outpatient therapy services for calendar years 2010 and 2011. The specific time period for the claims varied depending on the analysis. The 100 percent claims analyses had several purposes. One was to analyze case-mix adjustment for therapy payment using administrative data, including diagnosis-related groups and hierarchical coexisting conditions. A second purpose was to develop descriptive statistics on episodes of therapy care and to conduct simulations of various payment systems, including FFS, episode-based, and blended FFS/episode-based payment systems. Results for both sets of analyses are detailed in the *DOTPA Payment Alternatives Report* (Amico et al., 2014b) and are summarized in ***Section 4*** of this report.

[This page intentionally left blank.]

3. DOTPA MEASUREMENT REPORT

The overarching purpose of the Developing Outpatient Therapy Payment Alternatives (DOTPA) project is to identify characteristics of a possible alternative approach to reimbursement of outpatient therapy services. However, before the approach to reimbursement was assessed, data from the DOTPA Continuity Assessment Record and Evaluation (CARE) instruments were evaluated for use in defining various *functional status* measures. This evaluation took place because function status was believed to be a critical mechanism on which to establish payment.

Two different CARE item sets were identified to construct the functional status measures:

- CARE-C – designed for patients dwelling in the community; and
- CARE-F – designed for nursing facility residents and day rehabilitation program patients, who tend to have more functional limitations than community-dwelling patients.

The CARE items measure several types of functional ability such as physical mobility, cognition, language, and activities related to daily living and self-care. In general, the CARE item sets contain both a patient self-report and a clinician observed-report, from which functional subscales were created. The CARE-F instrument largely resembles the CARE-C instrument with two important distinctions: (1) a greater number of clinician-reported items were included in CARE-F to capture detail at lower levels of function; and (2) the patient self-report section was shorter in CARE-F in acknowledgment of the difficulty this population would have in completing the form.

The evaluation of the CARE-C and CARE-F item sets was conducted through a combination of consultation and analysis. The DOTPA project team consulted participating providers to gain insight into the utility of the CARE item sets in capturing the functional characteristics of Part B patients treated within the providers' facilities. Then, a technical expert panel (TEP) was convened with health care and measurement experts to obtain clinical and methodological feedback on initial psychometric analyses from the DOTPA data. After receiving TEP feedback on the importance of examining potential diagnosis group differences, the CARE item sets were further evaluated to establish psychometrically meaningful subscales from the clinically meaningful item sets. These psychometrically appropriate subscales were then used in subsequent payment analyses summarized in ***Section 4*** of this report. ***Tables 3-1*** through ***3-4*** contain detailed information on the CARE-C and CARE-F subscales, including item content and response option information (in footnotes), ordered first by patient self-report information and followed by subscales using clinician-reported information.

Detailed information on the results from measurement evaluation is contained within the *DOTPA Measurement Report* (Kline et al., 2014). The descriptive analyses of the patient self-report items are summarized in ***Section 3.1*** of this report, followed by the descriptive results for the clinician-reported items in ***Section 3.2***. The internal consistency of the items within the scales is described in ***Section 3.3***. Rasch techniques used to create the functional status subscales are described in ***Section 3.4***.

Table 3-1
CARE-C analytic subscale information: Patient self-report

Subscale	Item content
Basic Mobility[1,2]	How much help from another person do you currently need moving to and from a bed to a chair? How much difficulty do you currently have… a) Moving from sitting at the side of the bed to lying down on your back? b) Moving up in bed? c) Standing for at least 1 minute? d) Sitting down in an armless straight chair? e) Standing up from an armless straight chair? f) Getting into and out of a car/taxi? g) Walking around on one floor, taking into consideration thresholds, doors, furniture, and a variety of floor coverings? h) Going up and down a flight of stairs inside, using a handrail? i) Bending over from a standing position to pick up a piece of clothing from the floor without holding onto anything? j) Walking several blocks? k) Walking up and down steep unpaved inclines? l) Carrying something in both arms while climbing a flight of stairs?
Wheelchair Use[1]	Without help from another person, when you are using your wheelchair, how much difficulty do you currently have… a) Moving around within one room, including making turns in a wheelchair? b) Opening a door away from a wheelchair? c) Opening a door toward a wheelchair? d) Transferring between a wheelchair and other seating surfaces, such as a chair or bed? e) Propelling/driving a wheelchair several blocks?
Everyday Activities[1,2]	How much help do you currently need… a) Taking care of your personal grooming such as brushing teeth, combing hair, etc.? b) Bathing yourself? How much difficulty do you currently have… a) Picking up thin, flat objects from a table? b) Putting on and taking off a shirt or blouse? c) Putting on and taking off socks? d) Opening small containers like aspirin or vitamins?

(continued)

Table 3-1 (continued)
CARE-C analytic subscale information: Patient self-report

Subscale	Item content
Everyday Activities[1,2]	How much help do you currently need… a) Removing stiff plastic packaging using hands and scissors? b) Tying shoes? c) Unscrewing the lid off a previously unopened jar without using devices? d) Washing indoor windows? e) Lifting 25 pounds from the ground to a table? f) Cutting your toenails?
Life Skills[1]	How much difficulty do you currently have… a) Understanding instructions involving several steps? b) Following/understanding a 10- to 15-minute speech or presentation? c) Answering yes/no questions about basic needs? d) Making yourself understood to other people during ordinary conversations? e) Telling someone important information about yourself in case of emergency? f) Explaining how to do something involving several steps to another person? g) Reading and following complex instructions? h) Telling others your basic needs? i) Planning for and keeping appointments that are not part of your weekly routine? j) Reading simple material? k) Filling out a long form? l) Writing down a short message or note? m) Remembering where things were placed or put away? n) Keeping track of time? o) Putting together a shopping list of 10 to 15 items? p) Remembering a list of 4 or 5 errands without writing it down? q) Taking care of complicated tasks like managing a checking account or getting appliances fixed?
Participation[3]	Even with help or services, tell us how much you are limited in… a) Keeping your home clean and fixed up? b) Providing personal care to yourself? c) Getting groceries or other things for your home? How much are you currently limited in… a) Going to movies, plays, concerts, sporting events, museums, or similar activities?

NOTES:

CARE-C = Continuity Assessment Record and Evaluation form for patients living in the community.

[1] Response options: Unable, A Lot of Difficulty, A Little Difficulty, No Difficulty

[2] Response options: Unable, A Lot of Help Needed, A Little Help Needed, No Help Needed

[3] Response options: Extremely Limited, Very Much Limited, Somewhat Limited, A Little Limited, Not at All Limited

SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA), 2014.

Table 3-2
CARE-F analytic subscale information: Patient self-report

Subscale	Item content
Patient Ability[1,2]	How much difficulty do you currently have… a) Moving from sitting at the side of the bed to lying down on your back? b) Moving up in bed? c) Walking around on one floor, taking into consideration thresholds, doors, furniture, and a variety of floor coverings? d) Making yourself understood to other people during ordinary conversations? e) Telling others your basic needs? How much HELP from another person do you currently need… a) Moving to and from a toilet? b) Taking care of your personal grooming such as brushing teeth, combing hair, etc.?
Wheelchair Use[1]	Without help from another person, when you are using your wheelchair, how much difficulty do you currently have… a) Moving around within one room, including making turns in a wheelchair? b) Opening a door away from a wheelchair? c) Opening a door toward a wheelchair? d) Transferring between a wheelchair and other seating surfaces, such as a chair or bed?

NOTES:

CARE-F = Continuity Assessment Record and Evaluation form for patients who receive Part B therapy while residing in a nursing facility or receiving day rehabilitation services.

[1] Response options: Unable, A Lot of Difficulty, A Little Difficulty, No Difficulty

[2] Response options: Unable, A Lot of Help Needed, A Little Help Needed, No Help Needed

SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA), 2014.

Table 3-3
CARE-C analytic subscale information: Clinician observation

Subscale	Item content (category)
Self-Care[1]	▪ Oral hygiene ▪ Wash upper body ▪ Upper-body dressing ▪ Lower-body dressing ▪ Putting on/taking off footwear
Mobility[1]	▪ Sit to lying ▪ Roll left and right ▪ Lying to sitting on side of bed ▪ Sit to stand ▪ Chair/bed-to-chair transfer ▪ Picking up object while standing ▪ Walk 50 ft. with two turns ▪ Walking 10 ft. on uneven surfaces ▪ Four steps ▪ Twelve steps ▪ Wheel up and down ramp ▪ Walk 500 ft., walk 150 ft., walk 50 ft., walk in room once standing ▪ Wheel 500 ft., wheel 150 ft., wheel 50 ft., wheel in room once seated
Instrumental Activities of Daily Living[1]	▪ Medication management–oral ▪ Make a light meal ▪ Wipe down surface and clean the cloth
Cognition	
Problem Solving[2]	Simple Problems: Following basic schedules; requesting assistance; using a call bell; identifying basic wants/needs; preparing a simple cold meal without assistance. Complex Problems: Working on a computer; managing personal, medical, and financial affairs; preparing a complex hot meal; grocery shopping; route finding and map reading without assistance.
Memory[2]	Basic Information: Personal information (e.g., family members, biographical information, physical location); basic schedules; names of familiar staff; location of therapy area without assistance. Complex Information: Complex and novel information (e.g., carry out multiple-step activities, follow a plan); anticipate future events (e.g., keeping appointments) without assistance.

(continued)

Table 3-3 (continued)
CARE-C analytic subscale information: Clinician observation

Subscale	Item content (category)
Attention[2]	Simple Activities: Following simple directions; reading environmental signs or short newspaper/magazine/book passages; eating a meal; completing personal hygiene; dressing without assistance. Complex Activities: Watching a news program; reading a book; planning and preparing a meal; managing one's own medical, financial, and personal affairs without assistance.
Communication	
Spoken-Language Comprehension[2]	Basic Information: Simple 1-step directions; simple yes/no questions; simple words or short phrases; conversations about routine daily activities without assistance. Complex Information: Complex sentences, 2–3 step directions, 2–3 part messages, and a variety of complex topics without assistance.
Spoken-Language Expression[2]	Basic Information: Basic information regarding wants/needs or daily routines; using 1–2 words or short phrases without assistance. Complex Information: Thoughts/ideas using sentences; in conversations about routine daily activities or a variety of topics without assistance.
Motor Speech Production[2]	Intelligible in Short Utterances: Spontaneous production of automatic words, predictable single words, or short phrases in conversation without assistance. Intelligible in Longer Utterances: Spontaneous production of multisyllabic words in sentences without assistance.
Functional Voice[2]	Low Vocal Demand: Speaking softly; speaking in quiet environments; talking for short periods of time without assistance. High Vocal Demand: Speaking loudly; speaking in noisy environments; talking for extended periods of time without assistance.

NOTES:

CARE-C = Continuity Assessment Record and Evaluation form for patients living in the community.

[1] Response options: Dependent, Substantial/Maximal Assistance, Partial/Moderate Assistance, Supervising/Touching Assistance, Set-up/Clean-up Assistance, Independent

[2] Response options: Never/Rarely, Sometimes, Usually, Always

SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA), 2014.

Table 3-4
CARE-F analytic subscale information: Clinician observation

Subscale	Item content (category)
Self-Care[1]	▪ Eating ▪ Oral hygiene ▪ Toileting hygiene ▪ Wash upper body ▪ Shower/bathe self ▪ Upper-body dressing ▪ Lower-body dressing ▪ Putting on/taking off footwear
Mobility[1]	▪ Sit to lying ▪ Roll left and right ▪ Lying to sitting on side of bed ▪ Sit to stand ▪ Chair/bed-to-chair transfer ▪ Picking up object while standing ▪ Walk 50 ft. with two turns ▪ Walking 10 ft. on uneven surfaces ▪ One step (curb) ▪ Four steps ▪ Twelve steps ▪ Wheel up and down ramp ▪ Walk 150 ft., walk 100 ft., walk 50 ft., walk in room once standing ▪ Wheel 150 ft., wheel 100 ft., wheel 50 ft., wheel in room once seated
Instrumental Activities of Daily Living[1]	▪ Telephone–answering ▪ Telephone–placing call ▪ Medication management–oral ▪ Medication management–inhalant/mist ▪ Medication management–injectable ▪ Make a light meal ▪ Wipe down surface and clean the cloth ▪ Car transfer

(continued)

Table 3-4 (continued)
CARE-F analytic subscale information: Clinician observation

Subscale	Item content (category)
Cognition Problem Solving[2]	Simple Problems: Following basic schedules; requesting assistance; using a call bell; identifying basic wants/needs; preparing a simple cold meal without assistance. Complex Problems: Working on a computer; managing personal, medical, and financial affairs; preparing a complex hot meal; grocery shopping; route finding and map reading without assistance.
Memory[2]	Complex Information: Complex and novel information (e.g., carry out multiple-step activities, follow a plan); anticipate future events (e.g., keeping appointments) without assistance. Basic Information: Personal information (e.g., family members, biographical information, physical location); basic schedules; names of familiar staff; location of therapy area with assistance.
Attention[2]	Simple Activities: Following simple directions; reading environmental signs or short newspaper/magazine/book passages; eating a meal; completing personal hygiene; dressing without assistance. Complex Activities: Watching a news program; reading a book; planning and preparing a meal; managing one's own medical, financial, and personal affairs without assistance.
Communication	
Spoken-Language Comprehension[2]	Basic Information: Simple 1-step directions; simple yes/no questions; simple words or short phrases; conversations about routine daily activities without assistance. Complex Information: Complex sentences, 2–3 step directions, 2–3 part messages, and a variety of complex topics without assistance.
Spoken-Language Expression[2]	Basic Information: Basic information regarding wants/needs or daily routines; using 1–2 words or short phrases without assistance. Complex Information: Thoughts/ideas using sentences; in conversations about routine daily activities or a variety of topics without assistance.
Motor Speech Production[2]	Intelligible in Short Utterances: Spontaneous production of automatic words, predictable single words, or short phrases in conversation without assistance. Intelligible in Longer Utterances: Spontaneous production of multisyllabic words in sentences without assistance.

(continued)

Table 3-4 (continued)
CARE-F analytic subscale information: Clinician observation

Subscale	Item content (category)
Functional Voice[2]	Low Vocal Demand: Speaking softly; speaking in quiet environments; talking for short periods of time without assistance. High Vocal Demand: Speaking loudly; speaking in noisy environments; talking for extended periods of time without assistance.

NOTES:

CARE-F = Continuity Assessment Record and Evaluation form for patients who receive Part B therapy while residing in a nursing facility or receiving day rehabilitation services.

[1] Response options: Dependent, Substantial/Maximal Assistance, Partial/Moderate Assistance, Supervising/Touching Assistance, Set-up/Clean-up Assistance, Independent

[2] Response options: Never/Rarely, Sometimes, Usually, Always

SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA), 2014.

3.1 Overall Examination of Patient Self-Report Frequency Data

The psychometric evaluation began with examinations of the CARE-C and CARE-F patient-reported items to understand rates of missing data, reporter status (self, recorder, or proxy), community participation (e.g., the patient's difficulty with getting together with family/friends or going out of the house socially), and responses to a number of standalone CARE-C items that do not belong to subscales described in ***Table 3-1***. Standalone items comprised pain,[17] sadness, and confidence. The patient self-report section within the CARE instrument was at least partially completed for a majority of the sample cases (96 percent) across all provider types, especially within the CARE-C admission assessment. The patient was most likely to complete the self-report by him/herself in the outpatient setting, whereas patients in the nursing facility and day rehabilitation setting (CARE-F) were more likely to require at least some assistance to complete the assessment.

Key Findings: On CARE-C social community participation, overall, more than half of beneficiaries indicated that they had either few or no limitations in social activities (e.g., going out to visit family and friends). However, focusing on the discipline-specific results, only 24 and 29 percent of occupational therapy and speech-language pathology patients, respectively, indicated they had no difficulty continuing to interact socially compared with more than 50 percent of the physical therapy (PT) beneficiaries. In CARE-F community participation, nearly a third of day rehabilitation patients indicated no difficulty of continuing their normal habits of going out and interacting with others. Findings from the CARE-C admission data on pain show

---

[17] Pain was assessed by two different sources in the CARE item set. In CARE-C, the pain items were answered by the patients (patient self-report), whereas in CARE-F, the pain items were answered by a clinician (clinician-reported).

that among those who indicated any pain, more than three-quarters reported that their pain was greater than a 5 on a severity scale of 1 to 10 (10 being the most severe pain), about half indicated that they had pain that affected their sleep, and more than two-thirds reported having pain that limited their activities. Finally, an examination of the CARE-C sadness and confidence data indicated that more than half of patients reported no or very mild feelings of sadness, and three-quarters reported at least some confidence in their ability to complete activities that they like to do.

3.2 Overall Examination of Clinician Observation Frequency Data

Clinician-reported item rates included missing data, information on assessment method (directly observed, alternative, or not assessed), visual and hearing impairments, swallowing, pressure ulcers, incontinence, and, for CARE-F only, a set of pain items. To ascertain data on the clinician's assessment method, therapists were instructed to code the items on the CARE assessment based on direct observation of patient performance, and when not possible, an acceptable alternative source of information (e.g., observing the patient performing a related activity or movement, or a credible report from the patient of his or her ability to perform a particular activity). At the end of the assessment, therapists indicated the number of items in each section that were coded using the various methods and the number that were not assessed for the purpose of gaining a better understanding of standard evaluation methods. Data on the assessment method indicated that functional mobility items were much more likely than self-care or instrumental activities of daily living (IADL) items to be directly observed in the CARE-C settings. In addition, the CARE-F data indicate that nursing facility residents and the more medically complex patients are much more likely to have self-care items assessed through direct observation.

Key Findings: Overall, responses indicated that the majority of the DOTPA sample was functioning at the top of the ability range; this finding was consistent with feedback received during the follow-up conference call with participating DOTPA provider sites that used both the CARE-C and CARE-F. However, findings for the overall sample are heavily influenced by the size of the DOTPA CARE-C sample relative to the CARE-F. For example, specific findings for visual and hearing impairments, swallowing, pressure ulcers, and incontinence indicated that the CARE-F beneficiaries showed much higher rates of impairment than beneficiaries in the CARE-C sample. Specific findings on the pain items from the CARE-F admission assessments suggested that of those patients reporting pain, more than half said their pain was greater than a 5 on a severity scale of 1 to 10 (10 being the most severe pain). In addition, approximately half of those patients indicated that pain affected sleep, and approximately two-thirds had pain that affected their activities.

3.3 Internal Consistency (Reliability): CARE-C and CARE-F Patient-Reported and Clinician-Reported Items

Complementing the descriptive analyses, investigations of internal consistency were conducted on the patient self-report and clinician-reported subscales in CARE-C and CARE-F. The internal consistency analyses used a classical reliability measure to determine how well the items within a subscale interrelated. Results using the measure (Cronbach's alpha) confirmed that the item-level relationships in each subscale formed a cohesive unit (Cronbach's alpha statistics for all scales ranged between 0.73 and 0.99, all of which are above the expected minimum desirable value of 0.70).

The CARE instrument includes various item sets drawn from the Activity Measure for Post-Acute Care (AM-PAC) assessment. The AM-PAC item bank, from which the CARE self-report items were derived, has been validated in various care settings and diagnostic subpopulations (for example, see Andres, Haley, and Ni, 2003; Haley et al., 2006). Various forms of reliability—test-retest, subject-proxy (clinician or family member proxy), and internal consistency—have been established for the AM-PAC scales. These scales are referenced as Basic Mobility, Everyday Activities, and Life Skills in the CARE assessment. Specifically, high test-retest reliability (0.91–0.97) and subject-proxy reliability[18] (0.68–0.90) intra-class correlation coefficients (ICCs) have been reported for the three scales (Andres, Haley, and Ni, 2003). High setting–specific ICCs were reported for the Physical and Movement scale (0.91–0.93; called Basic Mobility in the CARE) and the Personal Care and Instrumental scale (0.82–0.93; called Everyday Activities in the CARE). Historical internal consistency reliability has also been reported to be high for the three AM-PAC scales, with Cronbach's alpha values of 0.90–0.95 for overall and diagnostic subgroups (Andres, Haley, and Ni, 2003; Haley et al., 2006).

3.4 Rasch Analysis

Rasch measurement analyses, an item-level analysis technique, supplemented the results from the overall descriptive examination and internal consistency analysis by evaluating the CARE-C and CARE-F subscales directly. Rasch analysis[19] provides insight into subscale performance through an examination of output from a rating scale analysis to evaluate the item hierarchy (ordering from easy to difficult) and clinical interpretation. Technically, in the analyses, we examined the extent to which the CARE item set's data fit the primary assumptions of the Rasch model[20] including item fit (i.e., the extent of agreement between observed and expected item responses),[21] subscale unidimensionality (i.e., the extent to which each scale measures a single construct), and item-level independence.

---

[18] The ICC of 0.68 is reported for the Applied Cognition scale (Andres, Haley, and Ni, 2003).

[19] Rasch analysis of a subscale estimates each item's level of difficulty and each person's ability measure based on the responses provided, using raw scores and response counts. Then, after the estimates are determined, the software can estimate expected values for every sample member. Because of the estimation structure within the Rasch analysis program (Winsteps; Linacre, 2012), missing data are not problematic for this analysis. The exception is an individual with completely missing data for the items used in the subscale; because no response exists on which to base an estimate, the ability measure is missing for this record.

[20] The primary assumptions of the Rasch model include the subscales measure a single dimension (unidimensionality), items within a subscale do not strongly relate to each other beyond the common dimension they share (local independence), and there is relative agreement between an item's expected response and the response observed (item fit).

[21] Item-fit statistics (i.e., indicators of misfit) are an indication of the extent of agreement between observed and expected item responses, and they reflect item-set effectiveness in describing a single domain or construct. If the item-fit statistics are outside a pre-specified acceptable range, person responses for this item are classified as unexpected and could generally suggest that the misfitting item is measuring a different domain (Bond and Fox, 2001).

3.4.1 Patient Self-Report: CARE-C and CARE-F

Rasch analysis of the CARE-C and CARE-F item sets at admission and discharge show that overall, the patient self-report subscales (***Tables 3-1*** and ***3-2***) were functioning adequately for the samples of interest. On the whole, the patient self-report subscales each represented a single construct. However, we made a few adjustments to the set of items included in the patient self-report subscales based on specific findings of the Rasch analysis. First, the CARE-C Life Skills subscale item *answering yes/no questions about basic needs* seemed to be "noisy" (i.e., there was a lack of agreement between expected and observed responses on this item), and we did not recommend retaining it for the subsequent resource utilization analyses (***Section 4***). Because the remaining items in this subscale also address communication- and cognition-related tasks, we recommend that this item undergo further evaluation in future research.

Second, the CARE-F self-report Patient Ability subscale items *making yourself understood* and *telling others your basic needs* did not function well, in that they did not fit the assumptions of the Rasch model. Also, these items did not make clinical sense in relation to other items in the subscale that concern motor functions (e.g., moving up in bed or taking care of personal grooming). Therefore, these items were not recommended for inclusion in the self-report Patient Ability subscale. However, *making yourself understood* and *telling others your basic needs* from the CARE-F item set should be assessed for their usefulness in analyzing resource use, but as single items and not part of a subscale. These recommendations are summarized in ***Table 3-5***.

3.4.2 Clinician Observed–Report: CARE-C Only

Rasch analysis of the clinician-reported subscales in the CARE-C showed person distributions that were skewed toward the high end of patient ability. This finding is consistent with the frequency analysis findings on other clinician-reported items and the participant focus group discussion mentioned previously. Further analysis of the CARE-C clinician-reported items indicated that the distributional skew persisted across differing body regions and primary reason for therapy groups. However, as expected, the extreme high-ability CARE-C beneficiaries did not drastically influence the operational definition (i.e., item hierarchy ordering from easy to difficult) of the CARE-C clinician-reported items. Therefore, these clinician-reported items measured the lower-ability individuals as intended. Future analyses with these data and the resulting conclusions should consider the impact of the ceiling effects when measurement involves populations similar to the DOTPA community sample.

Another aspect examined in the CARE-C clinician-reported subscales was their fit to the Rasch model, with the idea that items showing misfit may be modified to maintain usefulness in future analyses. First, the CARE-C Mobility distance items (e.g., *walking 50 feet*, *wheeling 50 feet*) showed some misfit to the Rasch model and were subsequently combined into two individual assistance level items—one for walkers, and one for wheelchair users. This item records the patient's need for assistance of another person, regardless of distance traveled.

Next, the item *pick up object*, which showed misfit in the analysis, should be limited to people who are standing when picking up the object. Otherwise, it is difficult to determine whether the data were collected as expected.

Finally, the clinician-reported subscale evaluation involved the potential inclusion of wheelchair use within the CARE-C Mobility scale. Modifications to the Mobility subscale initially looked at wheelchair use separately from the remaining mobility items. This separation was because the upper-body region, typically associated with self-care items, is used to propel a wheelchair. However, wheelchair use concerns the ability to move through space and belongs clinically with the rest of the mobility items. Although the wheelchair use items are clinically important in the assessment of mobility, they did show some misfit to the Rasch model. However, there was some clinical discussion with the TEP that this may be due to the differences between manual and motorized wheelchair use; therefore, the type of device should be collected in future work.

3.4.3 Clinician Observed–Report: CARE-F Only

CARE-F clinician-reported items did not exhibit ceiling effects and generally met the assumptions of the Rasch model, which includes the requirement of measuring a single construct. As with the CARE-C, the CARE-F Mobility items required some modifications. First, the distance items (i.e., within the CARE-F instrument, questions C13.a–C13.d for ambulatory patients and questions C14.a–C14.d for those who use a wheelchair) showed some misfit to the Rasch model. To remedy the misfit, the four walking distance items were recoded into a single variable for further analysis, and the same was implemented for the wheelchair use distance. These recoded items were eventually used in a single CARE-F Mobility subscale (see the ***Table 3-5*** discussion as follows), and this combined scale was used in the payment alternatives analysis (Pope et al., 2014).

In addition, analysis suggested that the bed mobility items were associated with a slightly different construct than those expressed in the “general” mobility subscale (e.g., ambulation, wheelchair use, transfers). The bed-related items, however, are important components of overall mobility functioning. Therefore, it makes sense clinically to keep the bed-related items with walking and wheelchair mobility. Another item of note, the item *pick up object* also showed misfit and should be limited to standing patients only.

Finally, after reconsidering the content of the IADL subscale, we determined that the *car transfer* item assessed mobility function. Therefore, *car transfer* fit better with the mobility items from a clinical standpoint, and the change was made after the psychometric analysis, but before the assessment of resource use discussed in ***Section 4***.

3.4.4 Clinician Observed–Report: Communication and Cognition

In terms of other constructs within the CARE item set, clinician-reported items assessing communication, problem solving, memory, and attention appeared to be functioning well. Although potential subscale aggregations (e.g., combining the subscales for problem solving, memory, and attention into a single cognition subscale) were discussed within the project team, the TEP members who convened in January 2013 did not recommend aggregations for this population. Therefore, we concluded that further analysis of resource utilization under DOTPA should include the three *Cognition* (problem solving, memory, and attention) and four *Communication* (spoken language expression and comprehension, motor speech production, and functional voice) subscales individually.

3.4.5 Additional Considerations

The items within each of the CARE-C and CARE-F patient self-report and clinician observed report subscales were found to be statistically similar during the Rasch analyses. For example, in the clinician observed Mobility subscale, *walking on uneven surfaces* and *walking four steps*, showed similarity in difficulty level (i.e., the level of effort for participants). Therefore, these items could be seen as statistically similar, although they are assessing different functional tasks. To some extent, those statistical similarities were found to be consistent across diagnosis groups. During the January 2013 meeting, TEP members recommended against removing specific items from the subscales; therefore, items showing similarity in difficulty level in the CARE-C and CARE-F patient self-report and clinician observed report subscales were used in the analyses of resource use discussed in ***Section 4***.

Several extensions of Rasch analysis were used to investigate the concordance between similar patient self-report and clinician-reported subscales. As part of this analysis, we considered whether items from both sources could be used within a single scale. As a result of those studies, we recommended that both the self-report and clinician observation item sets be included in the evaluation of resource utilization, but in separate (not combined) subscales. Both data sources have benefits and drawbacks, and if used alongside one another in future evaluations, they could provide a more robust look at patient functioning.

Table 3-5 provides a summary of revisions to the original subscales after we completed all the psychometric analyses, and it includes recommendations for future administrations of the instruments. The revisions were implemented in constructing the final patient self-report and clinician-reported functional ability scales before their inclusion in the resource utilization analyses summarized in ***Section 4***. The item sets used in the subscales are included in ***Appendix A*** and as Appendix D.2 in the *DOTPA Payment Alternatives Report* (Amico et al., 2014b). The Rasch scores for the finalized subscales are displayed in ***Appendix B***, along with a "raw" scale score calculated by adding the response value from the assessment for all items within the subscale. Note that the raw scale score was calculated for only those beneficiaries with responses to all items within the subscale.

Table 3-5
Overall item-level recommendations

Subscale	Recommendation
CARE-C Subscales	
Self-Report	
Basic Mobility	No changes
Wheelchair Use	No changes
Everyday Activities	No changes
Life Skills	Remove *answering yes/no questions about basic needs* responses from subscale data
Participation	No changes

(continued)

Table 3-5 (continued)
Overall item-level recommendations

Subscale	Recommendation
Clinician Observation	
Self-Care	No changes
Walking Mobility	Combine all distance items into a single assistance-level variable Restrict *picking up object* item to only the walking sample
Wheelchair Mobility	Combine with Mobility subscale data Combine all distance items into a single assistance-level variable
Instrumental Activities of Daily Living (IADL)	No changes
CARE-F Subscales	
Self-Report	
Patient Ability	Remove *making yourself understood to other people during ordinary conversations*, *telling others your basic needs*, and *taking care of your personal grooming such as brushing teeth, combing hair, etc.* responses from subscale data
Wheelchair Use	No changes
Clinician Observation	
Self-Care	No changes
Walking Mobility	Combine all distance items into a single assistance-level variable Restrict *picking up object* item to only the walking sample Include *car transfer* responses from the IADL subscale
Wheelchair Mobility	Combine with Mobility subscale data Combine all distance items into a single assistance-level variable
IADL	Remove *car transfer* responses from subscale data

NOTES:

CARE-C = Continuity Assessment Record and Evaluation form for patients living in the community; and CARE-F = CARE form for patients who receive Part B therapy while residing in a nursing facility or receiving day rehabilitation services.

SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA), 2014.

[This page intentionally left blank.]

4. DOTPA PAYMENT ALTERNATIVES REPORT

The purpose of the Developing Outpatient Therapy Payment Alternatives (DOTPA) project is to explore and identify characteristics of one or more approaches for reimbursing outpatient therapy under Medicare Part B that could serve as a refinement of or alternative to the current fee-for-service (FFS) plus annual expenditure cap system. Outpatient rehabilitation under Part B—physical therapy (PT), occupational therapy (OT), and speech-language pathology (SLP)—exhibits considerable variation in terms of patient case-mix characteristics, prior medical history and treatment, setting types, coding practices, clinical schools of thought, and the resulting treatment plans and utilization patterns.

The results from the DOTPA project, presented in this and the previous section, are exploratory in nature and serve as a starting point for a possible alternative payment system. The research presented in the *DOTPA Payment Alternatives Report* (Amico et al., 2014b) and summarized in this section does not aim to provide a complete or final model for a new payment system. Instead, selected characteristics associated with a possible alternative payment system are analyzed and discussed. Results are presented, including an analysis of patient case-mix not previously available, with the goal of enabling the Centers for Medicare & Medicaid Services (CMS) to make informed decisions about the payment alternatives and research agenda moving forward. Some of the current research is focused on specific alternatives of interest to CMS, as well as others identified as potential strong candidates by the DOTPA project team and members of the DOTPA technical expert panel meeting held in January 2013.

A major contribution of this research on therapy payment is the analysis of real-time beneficiary data collected with the Continuity Assessment Record and Evaluation (CARE)-C and CARE-F instruments (***Section 2.3***), along with characteristics of the participating DOTPA provider sites. The CARE-C and CARE-F instruments were administered at both admission and discharge to capture change in the beneficiary's condition after a course of therapy. In keeping with the goal to identify case-mix groups with similar resource utilization patterns used to predict needs for future beneficiaries, only the admission assessments are used in the payment analyses summarized as follows. The discharge assessment data are reserved for future research that additionally capture patient status changes from admission to discharge.

In addition to the DOTPA patient report/clinician assessment data, we analyze 100 percent of the Medicare Part B outpatient therapy claims from the period 2010–2012. The claims and other administrative data are analyzed both separately and in conjunction with the CARE-C and CARE-F data.

The following sections contain a summary of the key findings detailed in each section of the *DOTPA Payment Alternatives Report*, including a discussion of relevant data items, the methodology used in our analysis, and implications of the results. ***Section 4.8*** contains some concluding observations based on the results presented in ***Sections 4.1*** through ***4.7***.

4.1 Therapy Payment Issues and Alternatives

In this section, we review therapy payment issues and alternatives, beginning with a description of the current Medicare payment approach for outpatient therapy services. We outline three options for a revised payment system as follows. Several other payment options were also considered but are not addressed (see Section 2.4 in the *DOTPA Payment Alternatives Report*).

4.1.1 Current Payment Policy

An important aspect of the FFS system is an annual, discipline-specific cap on beneficiary utilization. This limitation covers services received in all outpatient therapy settings. PT and SLP are subject to a combined cap, whereas OT is subject to a separate cap. For calendar year 2013, each cap was set at $1,900 in allowed charges. The policy of the annual cap has undergone several recent revisions including extension to care received in hospital outpatient departments (originally exempt), and an additional manual medical review when allowed charges exceed $3,700 per year.

4.1.2 Problems with the Current Approach

Since 2006, Medicare expenditures for outpatient therapy have grown at an average annual rate of 8.1 percent. The FFS system currently in place is structured in a way that can generate high administrative burden and may incentivize high-service use and cost growth. The procedure codes used to identify the services provided are complex to report; yet, they yield very little meaningful information on the underlying needs of the beneficiary. Additionally, the therapy cap is a universal and soft limitation that is not sensitive to the needs of the patient. Exceptions allow high-cost cases to continue with documented need; 81 percent of patients never reach the cap and are unaffected by the exceptions process. Very little attention is paid to these lower-cost cases and whether the treatment is clinically appropriate. Finally, medical reviews of cases exceeding the cap are costly and burdensome for CMS and the Medicare Administrative Contractors, and complete enforcement of appropriate treatment is not feasible.

4.1.3 Three Options for a Revised Payment Approach

Refinement of Current Fee-for-Service System and Annual Cap

The first option is a refinement of the current FFS system and annual cap. This option has the benefit of an existing infrastructure and familiarity on the part of providers. Suggested refinements include increasing beneficiary cost sharing, risk adjustment of the cap amount, reducing provider payments above the cap, and moving toward completely separate or combined caps. Most of these options are focused on cost containment and may have a limited effect. Beneficiary cost sharing, for example, could be blunted because of the large majority of patients who have supplemental insurance plans and was not considered further. Characteristics of a new system with reduced provider payments above the cap were also not evaluated, owing to the possible implications for care of beneficiaries requiring extensive therapy. Although precise risk adjustment is necessary to accurately adjust the cap, this option was evaluated using the claims and assessment responses for the DOTPA sample.

Episode-Based Payment

A second option investigated is episode-based payment. This option bundles all services provided over a period of time into an episode of care; payment is made at the “episode” unit.

This technique has the advantages of both incentivizing efficiency and clinical face validity. Episode-based payment is also a helpful mechanism for future research involving outcomes, efficiency, and quality, all of which are difficult to measure under the current system. Episode-based payment presents challenges, however. By increasing efficiency, bundled payments can incentivize underprovision of appropriate care and can also put providers at greater financial risk. Further, defining episodes and appropriate payments in a population as clinically diverse as outpatient rehabilitation is complex. Precise case-mix adjustment is important to the success of any episode-based payment system. Analyses conducted in the DOTPA project included (1) an episode with no pre-specified length of time (referred to as a variable-length episode) with a 60-day clean period (i.e., a 60-day period with no prior therapy use), and (2) an annual period that covers either a calendar year or a 12-month period. Some patients will naturally be more complex and many also have truncated episodes from situations such as unplanned withdrawal from therapy or transfer to another provider. All of these situations must be accounted for to ensure that CMS is appropriately paying providers for services rendered.

Mixed FFS Payment

A third option, mixed FFS, is a hybrid of FFS and episode-based payment. In this approach, a smaller bundled payment, constituting a certain percentage of the total expected payment, is given to the provider, who retains the lump sum regardless of how much treatment is actually delivered. The bundled payment is distributed at the first visit along with a reduced FFS payment, and subsequent visits would be reimbursed only with reduced FFS payments. FFS payments would be reduced compared with current levels, thereby incentivizing less services use. This option recognizes the balance between the need to reduce the incentive for overprovision under FFS and the incentive for underprovision under an episode-based payment. It also manages incentives in a way that episode-based payment or FFS alone cannot by adjusting when the episode-based payment ends and the FFS payment begins. This policy could be employed either as the primary approach to reimbursement or as a transitional measure to help develop an episode-based payment system in the long run. A mixed FFS payment has the same level of complexity as an episode-based payment, but decreases the incentive for underprovision of care. However, the bundled payment would need to be risk adjusted, and there is incentive for providers to proliferate more episodes, because of the bundled payment at the front end of the episode.

4.2 DOTPA Analytic File Methodology

4.2.1 Overview of Claims-Based Variables

The dependent variable, Medicare-allowed charges, was extracted from claims data. The remaining explanatory variables were obtained from claims and enrollment administrative data. Explanatory variables from claims and administrative data include age, sex, original reason for entitlement (e.g., originally disabled), current reason for entitlement (e.g., currently disabled), dual-eligible status (Medicaid), end-stage renal disease (ESRD) status, State (used to create nine census divisions), diagnosis-related group (DRG), hierarchical condition categories (HCCs), hospital facility type, type of stay (short, long, skilled nursing facility), whether the patient’s inpatient stay (if any) included an outlier payment, and whether there were any PT, OT, or SLP charges during the hospitalization.

4.2.2 Overview of CARE Variables

Data collection on the CARE-C assessments was discipline specific (i.e., if a patient was treated by multiple disciplines, each discipline completed a separate CARE-C assessment form). Different subsets of variables from the CARE-C admission assessments were used in discipline-specific payment analysis, based on relevance of the variables for the respective discipline. Data collection on the CARE-F assessments was not discipline specific (i.e., if a patient was treated by multiple disciplines, all clinicians reported clinical assessment data on the same CARE-F form).

Given the small number of CARE-F day rehabilitation (day rehab) assessments and that day rehab is not a distinct Medicare provider type, only descriptive analyses of therapy expenditures are reported for the CARE-F day rehab sample. Risk-adjusted payment analyses using claims, enrollment, and CARE assessment data were conducted only for the CARE-C and CARE-F nursing facility samples.

4.2.3 Diagnosis and Primary Reason for Therapy Groups

The assessments collected data on primary diagnosis, secondary diagnosis, and primary reason for therapy. Primary reason for therapy included body functions, body structures, and activities and participation, defined based on the International Classification of Functioning, Disability and Health (World Health Organization, 2001). Given the extensive list of primary diagnoses, secondary diagnoses, and primary reason for therapy items, groupings of related items were created by the DOTPA project team for each of these variables for use in payment analysis. The intent was to define groups that had optimal within-group homogeneity in terms of both clinical characteristics and resource needs. For the CARE-C sample, although we initially examined the possibility of defining one set of groups for all disciplines, discipline-specific diagnosis groups were deemed necessary given the differences in scope of practice, patient populations, and patient problems addressed by the separate disciplines (see Table 3-8 in the *DOTPA Payment Alternatives Report*). Twenty-one mutually exclusive primary diagnosis groups were created for PT, and 9 were created for OT. Twenty-three secondary diagnosis groups were created for PT, and 11 were created for OT. The primary diagnosis groups were intended to reflect the underlying etiologic diagnosis for which the beneficiary needed PT or OT, whereas the secondary diagnoses were intended to represent coexisting conditions and comorbidities. For SLP, primary and secondary diagnoses were combined to create a set of four mutually exclusive *Impairment Diagnosis* groups and a set of four mutually exclusive *Medical Diagnosis* groups (see Appendix Tables 3-9a and 3-9b in the *DOTPA Payment Alternatives Report*). Discipline-specific groups were also created for body functions (8 for PT, 7 for OT, and 4 for SLP), body structures (16 for PT, 7 for OT, and 4 for SLP), and activities (4 for each discipline), all reported under primary reason for therapy.

Given CMS’s objective to analyze a combined expenditures cap for PT, OT, and SLP in only the CARE-F nursing facility setting, CARE-F nursing facility payment analysis was not discipline specific. Therefore, unlike the discipline-specific groups created for CARE-C, a single set of groups was created for the CARE-F assessments (see Table 3-10 in the *DOTPA Payment Alternatives Report*). Eight primary diagnosis groups were created, along with 18 secondary diagnosis groups, 10 body structure groups, 5 body function groups, and 4 activity groups.

4.2.4 Rasch Function Subscales

The assessments also collected information on various aspects of functional status through both patient self-report and clinician observation. Beneficiaries were asked to respond to a series of questions about their ability to perform various functional activities, whereas clinicians were asked to rate beneficiary performance on various functional activities based on observation. From these individual self-report and clinician observed items, clinically related items were grouped into function subscales using Rasch analysis, a statistical procedure to create items sets for measurement. A total of 15 function subscales were produced from the CARE-C assessments—5 self-report subscales and 10 clinician-observed subscales. The five self-report item sets include *Mobility*, *Wheelchair Use*, *Everyday Activities*, *Life Skills*, and *Participation*. The clinician-observed subscales include three distinct Physical Function subscales (*Self-Care*, *Mobility*, and *Instrumental Activities of Daily Living*), and seven distinct Cognition and Communication subscales (*Problem Solving*, *Memory*, *Attention*, *Spoken-Language Comprehension*, *Spoken-Language Expression*, *Motor Speech Production*, and *Functional Voice*). A total of 12 function subscales were produced from the CARE-F nursing facility assessments—2 from the Self-Report section of the instrument and 10 from the Clinician-Observed section. The two CARE-F self-report item sets include *Mobility*[22] and *Wheelchair Use.* The 10 clinician-observed subscales were named identically to the CARE-C clinician-observed subscales, although some items differed.

From these item sets, relevant subscales were used in discipline-specific CARE-C payment analyses and combined analyses for the CARE-F nursing facility sample. Rasch estimates ranged from 0 to 100, with 0 indicating lowest functional ability and 100 indicating the highest ability. Mean estimates on the Rasch function subscales were lowest for the CARE-F nursing facility sample and highest for the CARE-C sample; this finding is expected because the functional ability of nursing facility patients is lower in comparison with the community population. In the CARE-C sample, Rasch estimate distributions on the clinician observation subscales were skewed toward the high end of functional ability, with the distributional skew being stable across differing body regions and primary reason for therapy groups (i.e., beneficiaries are, relatively speaking, functionally independent with these activities overall and within subgroups defined by body region and by primary reason for therapy). In the CARE-F nursing facility sample, Rasch estimates were more evenly distributed.

4.2.5 Additional Items Used in Analyses

Individual items from the CARE-C admission assessments were also included in relevant discipline-specific analyses, including swallowing function indicators; number and time frame of prior surgeries, if any, for the presenting problem; duration of the patient's presenting problem;

---

[22] The CARE-F self-report Mobility subscale, termed Patient Ability in the *DOTPA Measurement Report* (Kline et al., 2014; see ***Table 3-2***), was modified based on recommendation and currently contains only items relevant to patient mobility. Therefore, the terminology was changed to more accurately reflect the current state of the item set used in the payment alternatives analyses.

mobility devices used; indicators of memory, cognition, and communication problems; indicators of sadness in the past 2 weeks; pain severity; and impact of pain on patients' activities and sleep in the past 2 days. In addition to these items, the CARE-F nursing facility analyses also included self-care and mobility functional abilities before onset of presenting condition, wheelchair use before onset of presenting condition, falls in the past year, moderate or severe cognitive impairment, evaluation or treatment for cognitive problems, respiratory impairment, endurance impairment, bladder/bowel impairment, ability to understand verbal content and express ideas/wants, inattention, disorganized thinking, and altered level of consciousness/alertness.

4.3 Therapy Annual Expenditures Cap: Alternative Policies and Risk Adjustment Using Administrative Data

Currently, Medicare expenditures on outpatient services are subject to an annual cap, using allowed charges, for PT and SLP combined, and a separate cap for OT. An exceptions process allows for Medicare coverage of expenditures above the caps (see ***Section 2***). The caps are the same for all beneficiaries, regardless of health status or age. In the following sections, we first present methods followed by descriptive results. Then, we simulate alternative policies for annual expenditure caps by individual or combined therapy disciplines. Finally, we analyze how the caps might be adjusted for beneficiary need for outpatient therapy services using only available Medicare administrative data. This process is referred to as "risk adjustment of the caps." We discuss multivariate regression models that predict annual therapy expenditures for both community and institutional residents using patient characteristics. For community residents, separate models are evaluated for PT, OT, and SLP. A summary of the methods used and analytic results obtained is provided as follows.

4.3.1 Methods

We analyze the complete set of therapy claims, 100 percent of all Medicare outpatient therapy claims, from calendar year 2011. The alternative therapy annual expenditures cap policies simulated are the following:

- current policy (equal caps for PT/SLP combined and for OT),
- equal discipline-specific caps, and
- a single combined cap for all three disciplines.

To make the three cap policies comparable, we require that they be budget neutral with respect to current policy. By budget neutral, we mean equal aggregate Medicare-allowed charges above and below the caps, assuming no behavioral response to changes in the caps by therapists or beneficiaries. We compare the cap policies on simulated 2011 dollar level of caps, number of Medicare therapy users above and below the cap, and average amount by which those beneficiaries above the cap exceed it.

The risk adjustment regression analyses predict annual expenditures for each therapy discipline separately. They include a *demographic-only model* that predicts annual therapy expenditures using demographic factors only (see Tables 4-5–4-7 in the *DOTPA Payment Alternatives Report*). In addition to the demographic model, we explore a prospective HCC model, a concurrent DRG model, and a concurrent inpatient utilization and prospective risk score model.

The *prospective HCC model* uses demographics and diagnoses from the prior year (HCCs) to predict current-year therapy expenditures (see Tables 4-5–4-7 in the *DOTPA Payment Alternatives Report*). The focus of the HCC model is on explaining therapy expenditures by the presence of chronic medical conditions that persist from the prior year into the current year, or prior-year acute conditions that have sequelae reaching into the current year. The HCC diagnoses are established from physician and hospital diagnoses recorded on Medicare claims over a 1-year period, and they are more likely to capture serious underlying chronic conditions than therapy claims, on which reporting of coexisting conditions is not required. The HCC model captures patient functional status only insofar as functional status is related to diagnosis—the HCC model does not directly measure functional status.

The *concurrent DRG model* includes demographics, as well as 49 clusters (groupings) of related DRGs that account for the majority of Medicare inpatient spending each year (see Tables 4-5–4-7 in the *DOTPA Payment Alternatives Report*). The DRGs are measured in the same year as therapy expenditures, so the DRG model is a “concurrent” model, as opposed to the HCC model, which is “prospective” because it uses prior-year diagnoses. DRGs measure diagnoses and procedures that result in hospitalizations in the current year. The focus of the DRG model is using current-year acute conditions and utilization to explain therapy expenditures.

The *concurrent inpatient utilization and prospective risk score model* includes demographics and variables measuring inpatient therapy utilization, type of inpatient facility in which inpatient stays occurred (including hospitals, skilled nursing facilities, inpatient rehabilitation facilities/units, and other non-acute, short-term hospital stays), outlier expenditure status of hospitalizations, and the HCC prospective risk score measuring prior-year diagnostic burden (see Tables 4-5–4-7 in the *DOTPA Payment Alternatives Report*). The model focuses on whether characteristics of inpatient utilization—in particular, inpatient therapy charges—predict outpatient therapy expenditures. This model also examines the relationship of patients’ overall disease burden as measured by the prospective HCC risk score to their therapy utilization.

4.3.2 Descriptive Results

There were 4,087,723 unique beneficiary users of outpatient therapy services (any discipline) in calendar year 2011. Of these, 3,655,812 beneficiaries used PT services (89 percent), 858,189 used OT services (21 percent), and 423,922 used SLP services (10 percent). OT and SLP users were somewhat older on average than PT users; the mean age ranged from 74 years for PT beneficiaries to 78 years for SLP beneficiaries.

In addition to age, we examined gender, ESRD status, currently entitled by disability (disabled) status, dual-eligibility (Medicaid) status, and long-term institutionalized (LTI) status of the beneficiaries across disciplines. PT, OT, and SLP users were predominately females. All of the disciplines had less than 2 percent ESRD patients, with the highest concentration being in OT. PT and OT had higher percentages of patients who were currently disabled (14 and 13 percent, respectively). OT and SLP had a higher share of dual-eligible beneficiaries (Medicaid) in comparison with PT. Thirty-nine percent of SLP users were dually eligible. LTI beneficiaries, those who were institutionalized for at least 1 month during the year, made up 10 percent of the overall therapy population. However, rates of institutional residence varied widely across the disciplines (8, 28, and 38 percent for PT, OT, and SLP, respectively). Last, 87 percent of therapy users were white, 8 percent were black, and the remaining 5 percent were from other racial or ethnic groups.

The average allowed charges for the 12-month period were $1,281 overall, $1,258 for PT, $1,365 for OT, and $1,306 for SLP. The median charges were $792, $814, $731, and $673 for the total, PT, OT, and SLP, respectively. Average allowed charges per therapy day were highest for SLP ($120), followed by OT ($92) and PT ($89). Annual therapy allowed charges varied 222-fold, from $36 at the first percentile to $7,762 at the 99th percentile. Annual allowed charge variation was driven mostly by variation in therapy days (75 at the 99th percentile to 1 at the first percentile), rather than by variation in allowed charges per therapy day ($195 at the 99th percentile to $29 at the first percentile).

The average patient received 14 distinct days of therapy during the course of a 12-month period, with a median of 9 therapy days; this pattern also existed across disciplines. Total calendar days (the number of calendar days from the first to the last therapy visit) averaged 68 days with a median of 34 days. OT and SLP appear to have more condensed and intense courses of therapy, averaging 58 and 49 calendar days, respectively, compared with 73 for PT. The increased intensity can be seen by the average number of therapy days per week, which ranged from 2.56 for PT to 4.12 for SLP.

4.3.3 Results on Simulated Annual Expenditure Caps

We simulate alternative non-risk-adjusted outpatient therapy annual expenditure cap policies. We compare equal discipline-specific caps and a single combined-discipline cap to the current policy of a combined PT/SLP cap and a separate OT cap. We find that a simulated 2011 budget-neutral, discipline-specific cap ($1,710) is lower than the actual 2011 therapy cap of $1,870, whereas a budget-neutral combined cap ($2,485) is higher. Discipline-specific caps are most favorable to beneficiaries needing a lot of services from all three therapy disciplines; a combined cap is most favorable to beneficiaries needing a lot of services from only one discipline. Along with these caps for all beneficiaries, we also simulate budget-neutral caps for community and institutional residents separately. The residence-specific caps follow the same general patterns as the all-beneficiary caps; however, the numerical values are different. For example, the budget-neutral combined cap for institutional residents is $2,959 instead of $2,485. We also analyzed the budget-neutral caps when using the $3,700 manual review threshold and found similar results.

4.3.4 Risk-Adjustment Results

Using four exploratory models, we investigate several sets of characteristics that could be used to risk-adjust the annual therapy expenditures cap: demographics, prior-year diagnoses (prospective HCCs), current-year hospitalizations (concurrent DRGs), and selected concurrent inpatient utilization variables (most notably inpatient therapy charges). A number of the individual demographic, diagnostic, procedure (surgical DRGs), and utilization variables predict statistically significant variations in annual therapy expenditures in the expected direction. These factors could be used to risk-adjust the annual cap.

However, the variables considered here explain little of the overall variation in annual outpatient therapy expenditures (see Tables 4-5–4-7 in the *DOTPA Payment Alternatives Report*); the percentage of variation explained (R^2) is uniformly under 10 percent. The concurrent DRG model improved on the demographics model to a lesser degree than did either the prospective HCC or the concurrent inpatient utilization models; this was particularly true for the OT and SLP models. Although combining the predictive factors of the four exploratory models in a single, discipline-specific model would undoubtedly raise the percentage of expenditure variation explained, it would still be relatively low.

As one step toward refining the regression prediction models, we estimated quantile regressions. The quantile regressions investigate the effect of risk factors on the therapy cap, simulated as the 77th to 81st percentile of expenditures, rather than on mean expenditures. The quantile regression results show a greater impact of institutional status, dual eligibility among the elderly, and age on the therapy cap than on mean expenditures. These results suggest that factors that tend to raise therapy expenditures—such as institutionalization, Medicaid enrollment, and older age—have a greater effect on the higher end of expenditure distribution than they do on average expenditures. Among groups such as the institutionalized, oldest old, and dually eligible, there is a higher proportion of individuals who have costly therapy. In risk-adjusting the therapy expenditures cap, the quantile regression suggests raising the cap more for persons with these characteristics than is suggested by the standard regression results.

4.3.5 Implications and Further Work

The low predictive power of these models raises the question of whether they would adequately risk-adjust the annual therapy expenditures cap. ***Section 4.4*** contains an examination of the results of adding functional status and other patient survey/clinician assessment variables to the set of explanatory information. The models investigated in this section are exploratory, and further exploration and development could be profitable before any changes in policy are made. For example, concurrent instead of prospective HCCs, and HCCs not used for Medicare Advantage payment but possibly relevant to predicting therapy expenditures, could be investigated. However, CMS does not routinely calculate these HCCs—unlike the prospective Medicare Advantage payment HCCs. In terms of utilization, we investigated only inpatient utilization and case-mix (DRG) variables. Further exploration could examine home health, skilled nursing facility, and hospital outpatient utilization and case-mix variables to determine if they can better predict outpatient therapy expenditures.

4.4 Risk-Adjustment of Annual Therapy Expenditure Cap Using Clinician Assessment and Patient Report Data

In this section, we discuss multivariate regression results that predict annual therapy expenditures. Building on the risk-adjusted annual cap using only administrative data presented earlier in ***Section 4***, we augment the analysis with functional-status data from the patients in the DOTPA sample. The DOTPA CARE-C and CARE-F instruments are used to collect data on patient function and other patient-reported and clinician assessment characteristics that may be useful in predicting beneficiary expenditures. After examination of a demographic-data-only model estimated from the DOTPA sample, we explore a *payment model* that includes other variables describing the patient's condition, as well as demographics. A third specification, the *comprehensive model*, includes additional covariates from the CARE-C or CARE-F instruments that may be predictive of expenditures, but not necessarily appropriate for a payment model because of their discretionary nature. These variables are included for exploratory purposes to understand non-payment factors that may drive expenditures.

4.4.1 Methods

The sample for the following analyses is any beneficiary who had a CARE admission assessment between March 1, 2011, and February 29, 2012. Annual expenditure data for the discipline matching the assessment was obtained for this 12-month period for each sample beneficiary who had a CARE-C assessment; all therapy expenditures were combined for CARE-F nursing facility patients. The demographic model is drawn solely from administrative data and includes age, gender, Medicaid status, ESRD status, and originally disabled status. The payment model includes the demographic variables and several additional variables from the CARE-C or CARE-F admission assessment such as diagnosis, function, structure, and activities groups. These variables were selected because they are related to clinical measures of patient complexity that are expected to be associated with expenditures. Excluded from this model are items that may also be associated with expenditures but may be more discretionary and therefore inappropriate for use in a payment model. The final model, the *comprehensive model*, is composed of the same items that are in the demographic and payment models but includes additional variables from the CARE analysis (e.g., depression indicators, regional adjusters, mobility devices, facility type) that may be predictive of expenditures. These additional variables are not necessarily appropriate for use in a payment model because of their more discretionary, or gameable, nature.

4.4.2 Descriptive Results

Overall, CARE-C had 4,842 unique beneficiaries and 4,856 assessments included in the 12-month time frame; 14 beneficiaries had assessments in multiple disciplines. Of the total assessments, 4,210 were PT assessments, 461 were OT assessments, and 185 were SLP assessments.

There was not a large difference in age between the disciplines; the mean age ranged from 72 to 73 years. On average, PT and OT patients were mostly females, whereas SLP patients were predominantly males. All of the disciplines have between 0 and 1 percent ESRD patients. OT and SLP have higher percentages of patients who are currently disabled. OT and SLP also have a higher share of dually eligible beneficiaries (Medicaid) in comparison with PT. Finally, 90 percent of the beneficiaries were white, 7 percent were black, and the remaining 3 percent were other race and ethnicity groups.

For the CARE-C population, the average allowed charges for the 12-month period were $1,493 overall, $1,488 for PT, $1,461 for OT, and $1,665 for SLP. Average allowed charges per therapy day were highest for SLP ($109), followed by OT ($98) and PT ($94). Annual therapy allowed charges varied 87-fold, from $82 at the first percentile to $7,173 at the 99th percentile.

Annual allowed charge variation was driven mostly by variation in therapy days (66 at the 99th percentile to 1 at the first percentile), rather than by variation in allowed charges per therapy day ($161 at the 99th percentile to $51 at the first percentile).

The average patient received 15 distinct days of therapy during the course of a 12-month period with a median value of 12 therapy days; this pattern also existed for all of the disciplines. Total calendar days averaged 95 days with a median of 52 days. OT and SLP appear to have more condensed and intense courses of therapy, averaging 75 and 80 calendar days, respectively, compared with 97 for PT.

CARE-F, on the other hand, had 519 assessments and 519 unique beneficiaries in the nursing facility population. Additionally, 169 beneficiaries had a unique assessment in the day rehab population. The average nursing home patient was 80 years old, whereas the average day rehab patient was 73 years old.

Nursing facility beneficiaries were composed of a much higher proportion of women (69 percent) than the day rehabilitation sample (49 percent). The nursing home sample had, on average, a higher proportion of ESRD (3 percent), disabled (10 percent), Medicaid (62 percent), and LTI (71 percent) patients than the day rehab sample. There were 1 percent ESRD, 10 percent disabled, 9 percent Medicaid, and no LTI day rehab patients. The nursing facility sample has a much higher composition of Medicaid and LTI patients. Day rehab patients were more likely to be white (91 percent) than nursing facility patients (86 percent), and 11 percent of nursing facility patients were black as compared with 2 percent of day rehab patients.

For the CARE-F population, the average allowed charges for the 12-month period were $5,390 for the nursing population and $4,521 for the day rehab population. Average allowed charges per therapy day were highest for day rehab patients ($147) and lower for nursing facility patients ($103). Annual therapy allowed charges varied 139-fold, from $231 at the first percentile to $31,952 at the 99th percentile for nursing facility patients and 47-fold for day rehab from $337 at the first percentile to $15,945 at the 99th percentile. For nursing facility patients, annual allowed charges were driven mostly by variation in therapy days (207 at the 99th percentile to 2 at the first percentile), rather than by variation in allowed charges per therapy day ($227 at the 99th percentile to $45 at the first percentile). Day rehab beneficiaries used fewer therapy days at the extreme (96) and had higher overall charges per therapy day at the 99th percentile ($328).

The average nursing facility patient received 48 distinct days of therapy during the course of a 12-month period with a median value of 31 therapy days. Day rehab patients received an average of 25 therapy days and a median value of 20 therapy days. Total calendar days averaged 152 with a median of 142 days for nursing facility patients and an average of 102 days with a median of 62 days for day rehab patients. The average number of therapy days per week is higher for nursing facility patients (2.85) than for day rehab patients (2.37).

4.4.3 Risk Adjustment Results: CARE-C Community Residents

Payment models for the prediction of therapy resources have relatively low levels of explanatory power. The OT payment model has the greatest explanatory power, explaining 22 percent of the overall variation in annual expenditures compared with 9 percent for PT and 11 percent for SLP (see Tables 5-3, 5-6, and 5-7 in the *DOTPA Payment Alternatives Report*). A substantial fraction of the explanatory power of the OT model relates to demographics, which alone explain 10 percent of OT expenditures. In contrast, demographics explain almost none of the PT and SLP expenditure variation in these samples. The payment variables do add substantial explanatory power to demographics for all three therapy disciplines, but even with the payment factors, only a small proportion of variation in annual therapy expenditures is explained. The additional factors in the comprehensive model add a substantial amount of explanatory power to the payment model, but these variables may not be appropriate for a payment model.

For PT annual expenditures, we find that primary diagnoses explain some variation in expenditures (3.6 percent) (see Table 5-4 in the *DOTPA Payment Alternatives Report*). Clinician-observed mobility adds an additional 2 percentage points of explanatory power. Adding all four Rasch function scales (Clinician-Observed mobility and three Self-Report scales) raises the percentage of variation explained to 6.2 percent. The clinician-observed and self-report Rasch function scales each add explanatory power to the model, but not at the same level as when all four scales are included simultaneously. This argues for collecting both the clinician-observed and self-report items if the marginal gain is greater than the additional cost of data collection. Demographics, primary diagnosis group, and four Rasch function scales explain about two-thirds of the expenditure variation that the full payment model explains (6.2 percent versus 9.3 percent). Demographics, primary diagnosis, and the Clinician-Observed mobility scale have about three-fifths of the explanatory power of the full payment model (5.4 percent versus 9.3 percent).

The Rasch function subscales almost always predict that higher function is associated with lower therapy expenditures, all other factors equal. This is as expected. However, the individual Rasch function scales are not statistically significant predictors of annual OT or SLP expenditures except for one anomalously signed OT coefficient. Two of the four Rasch function subscales included in the PT model—*self-reported participation* and *self-reported mobility*—are statistically significant predictors of annual PT expenditures. Their regression coefficients indicate that the difference between the highest ability (score = 100) to lowest ability (score = 0) is associated with a $401 and $290 increase in expenditures for each scale, respectively.

The *self-reported participation* Rasch subscale is the single payment variable that explains the most PT expenditure variation according to a forward stepwise regression, followed by *clinician-observed mobility* and a *diagnosis of joint replacement*. These three variables together account for more than half of the explanatory power of the PT payment model (5.2 percentage points of 9.3 percent). The payment model variables capture only 58 percent of the expenditure variation explained by the comprehensive model (adjusted R^2 of 9.3 percent versus 16.1 percent), indicating that factors that we suggest may not be suitable for payment have a strong impact on therapy expenditures.

A primary medical diagnosis of *joint replacement* for PT is a positive and statistically significant predictor of annual therapy expenditures. For example, a positive coefficient here indicates that a person who had a joint replacement would have higher annual therapy expenditures than a person who did not have a joint replacement, all other variables equal.

However, *joint replacement* may be capturing expenditures from a period before the procedure, as well as after the procedure. Primary diagnoses of *vertigo* and *pain* are negative predictors of PT expenditures. A primary diagnosis of *stroke* is a positive predictor of OT expenditures in the comprehensive model but not in the payment model. Several secondary diagnoses and reason for therapy body structures and functions are positive and negative predictors of PT and OT expenditures.

Using a *wheelchair/scooter full time* is a strong predictor of both PT and OT expenditures, but it is probably not an appropriate payment variable because providers may encourage increased wheelchair use because of payment incentives. Using a *wheelchair/scooter part time* is not associated with either PT or OT expenditures. Using a *walker* predicts PT expenditures, and using a *cane/crutch* predicts OT expenditures.

The SLP models are highly problematic because of small sample size—only 124 observations were available for the payment and comprehensive models. It is difficult to draw any conclusions from the SLP models because of the high likelihood of significant overfitting, collinearity, and outlier influence.[23]

The addition of provider identifiers significantly increases the explanatory power of the models. The annual payment model R^2 values increase from 9 to 32 percent, and the comprehensive model increases from 16 to 33 percent. The sizable increase in R^2 may be due to provider practice patterns, unmeasured case-mix characteristics, or some other unobserved factors that are correlated with the provider. The comprehensive model holds provider type (e.g., private practice) constant; thus, the effect is not due to provider type in the comprehensive model and may be due to individual provider practice.

4.4.4 Risk-Adjusted Cap Analysis: CARE-C Physical Therapy

Risk adjustment of the therapy cap is considered an option for an alternative payment system as a possible way to equitably distribute resources based on patient need. To this end, a risk adjustment analysis was conducted on the largest DOTPA sample—community-based beneficiaries receiving physical therapy.

For each of the 4,210 individual PT patients, we predicted their annual expenditures using their demographic and CARE assessment information multiplied by the regression coefficients and compared the predicted expenditures to their actual annual expenditures. Results show that the percentage of beneficiaries exceeding the cap in diagnosis groups with the highest percentage of beneficiaries exceeding the non-risk-adjusted cap—*parkinson's and other progressive neurological diagnoses*, *multiple major etiologies*, *joint replacement*, *unspecified and miscellaneous neurological diagnoses*, *and stroke*—is reduced by approximately 25 to 30 percent under the risk-adjusted cap. On the other hand, for the five diagnosis groups with the lowest percentage of beneficiaries exceeding the cap—*vertigo*, *pain*, *genitourinary disorders*, *no*

---

[23] Overfitting occurs with the use of model covariates associated with random error (i.e., noise in the model) instead of the dependent variable of interest. Collinearity occurs when two explanatory variables are highly correlated so that the independent contribution of each independent variable to the variation in the dependent variable cannot be determined.

major multiple etiologies, *and sprain/strain*—the percentage of beneficiaries exceeding the risk-adjusted cap increases by 160 to 450 percent. This is because these beneficiaries have a lower cap after risk adjustment than before risk adjustment. We also found that beneficiaries with the lowest functional scores exceeded the cap less frequently under the risk-adjusted cap. The percentage of cap exceeders in the highest functioning group increased from 63 percent of all cap exceeders to 72 percent under the risk-adjusted cap. Manual review policies based on the annual cap could be targeted at those high-cost beneficiaries who are not predicted to have high costs, using the characteristics measured in the risk adjustment model (e.g., primary diagnosis, Rasch mobility measure, and other characteristics used in the risk adjustment model). These policies would allow manual review to be more targeted, focusing primarily on those cases that have unexpectedly high costs.

4.4.5 Results of Risk-Adjustment: CARE-F Nursing Residents

When the basic payment variables are added to the demographics, the R^2 increases from explaining 4 percent of the variation in expenditures to 14 percent (8 percent adjusted) (see Table 5-10 in the *DOTPA Payment Alternatives Report*). The basic payment model includes demographics, primary diagnosis groups, Rasch clinician-observed mobility and self-care categorical variables, and indicators for diet modification, severe cognitive impairment, and verbal ability. This limited set of variables is intended to reduce the collinearity that is present in the full payment model because of the overlap between diagnosis, function, structure, and activities groups. In the basic payment model, the primary diagnosis groups are insignificant except the group that included no primary diagnosis coded; this group included only 20 patients. The Rasch mobility and self-care items were not significant; however, the self-care items increase monotonically, moving from the most functional to the least functional, which is as expected. The additional indicators are not significant.

When the additional payment variables are added to the model, the explanatory power increases from 14 percent to 36 percent (adjusted 25 percent). In this model, the only Rasch function scale that is significant is the self-reported mobility scale. For the ranges of score from 30 to 60, annual therapy expenditures increase by $1,664, and for the range from 60 to 90, annual expenditures increase by $2,216 as opposed to those with the highest scores (90 or higher). Annual expenditures are not affected by prior hospitalization in the past 2 months. Secondary diagnoses of *osteoporosis* or *hypertension* are important predictors of expenditures. In addition, *mental functions* and *motor functions* are important functional predictors of expenditures, and *shoulder/arm/elbow* and *knee* are important positive predictors based on the body structure involved.

An additional analysis uses a forward stepwise regression model and shows the marginal improvement in explanatory power as each variable is added to the model (see Table 5-11 in the *DOTPA Payment Alternatives Report*). The first three variables—daily activities, mobility, and other body functions—account for 41 percent of the overall explanatory power. Many of the variables that increased the explanatory power were poorly defined groups, such as *no primary diagnosis*, *activities not reported*, or *general/no specific body locations*. These may indicate beneficiary frailty, but it cannot be determined why they are important predictors of therapy expenditures without an expansion of the CARE instrument to collect more detailed information.

The final comprehensive model includes additional variables as described in Section 5.2.3 of the *DOTPA Payment Alternatives Report*. The addition of these variables increases the explanatory power of the model to 51 percent (adjusted 37 percent). Of the additional variables added to the model, the *patient's prior function in self-care being limited*, *having memory difficulty, and pain affecting activities* were all strongly significant positive predictors of expenditures. On the other hand, *prior wheelchair function being limited* and *rarely expressing ideas* were strong negative predictors of expenditures for the CARE-F patient population.

4.5 Episodes Defined Using Administrative Data: Descriptive Analysis and Payment Simulations

In this section, we describe the characteristics of outpatient therapy episodes of care using Medicare administrative data for all episodes. Separate episodes are evaluated for PT, OT, and SLP. We also use the episode data to conduct an exploratory simulation of non-risk-adjusted episode-based payment for initial fixed-length episodes of PT treatment during a year. The simulations range from full episode-based payment to various blends of FFS and lump-sum episode-based payment, referred to as mixed payment. The episode-based payment simulations are not intended to represent results from a fully developed episode-based payment system. Many additional elements—including risk adjustment—would need to be incorporated into a complete payment system. Rather, the simulations provide an exploratory look at the implications of full episode-based payment and blends of FFS and episode-based payment versus current FFS payment for initial episodes of PT treatment.

4.5.1 Methods: Episode Definition, Sample, and Data for Descriptive Analysis

To descriptively analyze episodes of therapy care, we rely on a *variable-length* episode definition. A variable-length episode definition does not pre-specify the length of episodes but organizes services into episodes based on the time pattern of therapy service utilization. Variable-length episodes rely on *clean periods* of no therapy utilization to define the beginning and end of the episode. We use 60-day initiating and terminating clean periods for our analysis. A new episode begins with a therapy service that is preceded by at least 60 days without any therapy claims in a discipline. An episode ends with a service that is followed by 60 days with no discipline-specific therapy service use.

We analyze all episodes of Medicare outpatient therapy care that began in CY 2010. To ensure a 60-day clean period before the start of 2010 episodes, we examined therapy claims starting in November 2009. We allow a 12-month run-out period for each episode from its start. If there is not a 60-day clean period by 12 months from the start of the episode, we censor (end) the episode at a 12-month length. Thus, our episodes may have lengths from 1 day to 1 year (12 months). Less than one-tenth of 1 percent of episodes is censored (see further details in Section 6 of the *DOTPA Payment Alternatives Report*). To allow the 12-month run-out for all episodes beginning in calendar year 2010, we examine claims through the end of 2011.

Episodes are specific to each of the three therapy disciplines. A beneficiary may have only one episode in a given discipline at a time, but he or she may have multiple and overlapping concurrent episodes across multiple therapy disciplines. A beneficiary may have multiple sample episodes in a single therapy discipline, as long as the episodes are separated by a 60-day clean period specific to that discipline. Episode services cover outpatient therapy services in all settings, including those billed through carrier claims or outpatient facility claims. For an episode to be included in the analysis file, the beneficiary must have been continually enrolled in Medicare Part B FFS, and Medicare must have been the primary payer for all months covered by the episode. Episodes are constructed based on 100 percent Medicare claims and enrollment data within the specified time period; thus, they include the universe of episodes satisfying the sample restrictions.

4.5.2 Methods: Episode-Based Payment Simulations

We simulate three types of episode-based payment: FFS (the current system), pure episode-based payment, and mixed (blended) FFS and episode-based payment. FFS payment for each episode is determined by summing the paid amounts on the Medicare claims for each therapy service assigned to an episode (see Tables 6-4–6-9c in the *DOTPA Payment Alternatives Report*). Pure (flat lump sum) episode-based payment is simulated as the mean of FFS payments for all qualifying episodes. The mixed payment is a blend of lump-sum payment per episode and a reduced percentage of FFS payment.

For the payment simulation, we limit the sample of beneficiaries to community residents who received PT beginning in 2010. We simulate episode-based payments only for a beneficiary's initial PT episode beginning in 2010. The start date of a beneficiary's initial episode is defined by a 60-day prior clean period, and the episode's end date is defined by a fixed period of 30, 60, or 90 calendar days from the start date. Qualifying therapy services falling within this time frame are assigned to the episode. Our simulation is thus limited in scope, but it explores the implications of episode-based payment for a sample of episodes that includes most community resident PT episodes beginning during a year.

4.5.3 Results: Descriptive Analysis of Episodes

Descriptive results are presented for all episodes (community or otherwise) initiating in calendar year 2010. The average age of beneficiaries with Part B therapy episodes beginning in 2010 is 74, with the average PT user's age being slightly younger and the average OT and SLP user's age being slightly older. Beneficiaries with therapy episodes are nearly two-thirds female, approximately 1 percent are entitled by ESRD, and slightly more than 14 percent are currently entitled by disability. Twenty percent of therapy episode users are Medicaid dual eligible individuals, ranging from 17 percent of PT users to 34 percent of SLP users. Ten percent of therapy episode users were classified as LTI residents in the same month that they initiated therapy, ranging from 6 percent of PT users to 31 percent of SLP users. Nearly double the proportion of beneficiaries with a high-cost episode resided in a nursing facility compared with beneficiaries with any therapy episode.

Outpatient therapy episodes beginning in 2010 have a mean allowed charge of $1,206. The mean allowed charge is similar for therapy discipline–specific episodes of PT, OT, and SLP. For all episodes, the distribution of episode total allowed charges ranges from $29 at the first percentile to $7,351 at the 99th percentile. The discipline-specific episodes show a similar range in total allowed charges.

On average, episodes of outpatient therapy last 42 calendar days, with a range from 1 day at the first percentile to 253 days at the 99th percentile (we censored episodes at a length of 12 months or 365 days). PT episodes last the longest, averaging 46 days; OT episodes are shorter, averaging 35 days; and SLP episodes are the shortest, averaging 27 days. At least 25 percent of SLP episodes last only 1 day, which may be partly due to the high prevalence of swallowing tests, which are a specialty of SLP clinicians. Median episode lengths are considerably shorter than mean lengths, indicating an episode-length distribution that is skewed to the right (i.e., most episodes are shorter than the average). A 60-day, fixed-length payment episode period would encompass about three-quarters of all PT episodes, well over three-quarters of all OT episodes and of all SLP episodes. A 30-day, fixed-length payment episode period would encompass about half of all PT episodes, more than half of OT episodes, and nearly three-quarters of SLP episodes.

Therapy days per outpatient therapy episode average 11–13, with a median of 9 (total, PT, and OT) or 5 (SLP). Allowed charges per therapy day are highest for SLP episodes, at an average of $124 per episode, and they average $97 for OT and $93 for PT. For all episodes, allowed charges per therapy day range from $24 at the first percentile to $206 at the 99th percentile, with a median of $93. The average intensity of therapy, measured as therapy days per week, is highest for SLP (4.54), intermediate for OT (3.88), and lowest for PT (2.91).

4.5.4 Results: Episode-Based Payment Simulations

We conduct exploratory simulations of several non-risk-adjusted episode-based payment variations for initial PT fixed-length episodes beginning in 2010. This analysis does not incorporate risk adjustment and is performed as groundwork to inform future work. The bundling or averaging inherent in episode-based payments means that some episodes are going to be paid less than current FFS payments, and some episodes will be paid more. Long-duration, higher-cost episodes will be paid less, and short, lower-cost episodes will be paid more under an episode-based payment policy. Given the substantial variation in therapy episode expenditures, pure lump-sum episode-based payment would result in substantial changes in payment for many episodes.

We examine two parameters of episode-based payment that could be used to lessen the changes from implementing episode-based payment. One parameter of fixed-length episode-based payment is the length of the episode. A shorter episode—30 days rather than 60 or 90 days—will result in less change from the current FFS payment. Another parameter of episode-based payment is the proportion of FFS versus lump-sum episode-based payment. Mixed episode-based payment (i.e., a combination of FFS and flat-rate payment) can achieve any blend from pure FFS payment to pure lump-sum episode-based payment. Higher blends of FFS payment result in less payment change from FFS. Choosing the best blend of FFS and flat-rate episode-based payment involves trading off the strength of incentives for efficiency and cost control (better with higher blends of flat-rate payment) against (1) lower incentives to entirely avoid or to undertreat beneficiaries needing more therapy, (2) lower incentives to create more episodes, and (3) less financial risk for therapy providers (better with higher blends of FFS payment).

4.5.5 Implications

There is tremendous variation in the length and cost of outpatient therapy episodes (and to a lesser extent in cost per therapy day) that is mostly not explained by type of therapy discipline involved. Total episode allowed charges vary from $29 at the first percentile to $7,351 at the 99th percentile, a ratio of 253 to 1. Explaining this huge cost variation, even with detailed beneficiary and clinical factors, is a difficult challenge. Given that a sizeable portion of cost variation is unlikely to be explained by available beneficiary and clinical factors, the implications of this variation for alternative payment systems need to be carefully considered. A desirable payment system will establish incentives to reduce any inefficiencies that may be reflected in this variation, but it would also need to ensure access to beneficiaries in the extreme part of the distribution who need high levels of service.

4.6 Risk Adjustment of Episode-Based Payments Using Clinician Observed Report and Patient Self-Report Data

In this section, we discuss multivariate regression results that predict episode-based therapy expenditures. Risk adjustment of episode-based payments is accomplished using a regression model that relates patient characteristics to episode expenditures. The descriptive episode analysis in Section 6 of the *DOTPA Payment Alternatives Report* used only administrative data. Section 7 of the *DOTPA Payment Alternatives Report* uses functional and other clinician-assessment and patient self-report data on the 6,490 patients in the CARE-C and CARE-F sample and estimates multivariate regression models to predict therapy episode expenditures. As discussed previously, the CARE-C and CARE-F instruments provide data on patient function that may be useful in predicting beneficiary expenditures. The analysis includes a demographic-only model to examine the impact of demographics alone; this analysis also serves as a baseline estimate. In addition to the demographic model, we explore a payment model that includes demographics, as well as other variables that may be relevant to a payment model. The *comprehensive model* includes additional covariates from the CARE-C or CARE-F assessment that may be predictive of expenditures but may not be appropriate for a payment model because of their discretionary nature.

4.6.1 Methods

The sample for the analyses is any beneficiary who had a CARE-C or CARE-F admission assessment between March 1, 2011, and June 30, 2012, which encompasses the full set of CARE-C and CARE-F admission assessments. Expenditure data for the discipline matching the assessment were obtained from the date of the CARE assessment until a 60-day period with no claims in that discipline; all expenditures were combined for CARE-F beneficiaries. Episodes were initiated by a CARE assessment; therefore, we examined only the episode associated with the assessment and did not explore subsequent episodes. All CARE episodes have at least a 6-month run out (June 30 to December 31, 2012) from their admission assessment, with most having a considerably longer run-out period.

The demographic model is drawn solely from administrative data and includes age, gender, Medicaid status, ESRD status, and originally disabled status. The payment model includes the demographic variables and several additional variables from the CARE-C or

CARE-F admission assessment such as diagnosis, function, structure, and activities groups. These variables were selected because they are measures of patient complexity that are expected to be associated with expenditures. Excluded from this model are items that may also be associated with expenditures but are considered less clinically necessary (discretionary) and therefore inappropriate for use in a payment model. The final model, the *comprehensive model*, is composed of the same items that are in the demographic and payment models but includes additional variables from the CARE-C or CARE-F analysis that may be predictive of expenditures. These additional variables are not necessarily appropriate to be used in a payment model because of their more discretionary, or gameable nature.

4.6.2 Descriptive Results

Overall, CARE-C had 5,476 unique beneficiaries, and 5,545 CARE-C admission assessments are included in the episode definition. Of these, 4,825 had PT assessments, 533 had OT assessments, and 187 received SLP assessments.

There is not a large difference in age between disciplines; the mean age ranged from 71 to 73 years. In addition to age, we examined gender, ESRD status, originally disabled status, and Medicaid status of the beneficiaries across disciplines. On average, the majority of PT and OT assessments were conducted on females, whereas SLP assessments were conducted on a majority of males. PT and SLP had less than 1 percent ESRD patients, and ESRD beneficiaries slightly exceeded 1 percent of the total beneficiaries receiving OT. OT and SLP had higher percentages of beneficiaries who entered Medicare as disabled than PT. OT and SLP also had a higher share of dually eligible Medicaid beneficiaries in comparison with PT. Last, 91 percent of the beneficiaries were white, 7 percent were black, and the remaining 2 percent were other races. OT and SLP had slightly more black beneficiaries.

For the CARE-C sample, the average allowed charges for the episode were $1,350 overall, $1,335 for PT, $1,320 for OT, and $1,825 for SLP. Average allowed charges per therapy day were highest for SLP ($107) and lower for OT ($98) and PT ($94). Overall, episode therapy allowed charges varied 78-fold, from $85 at the first percentile to $6,701 at the 99th percentile. Episode allowed-charge variation was driven mostly by variation in therapy days (60 at the 99th percentile to 1 at the first percentile), rather than by variation in allowed charges per therapy day ($160 at the 99th percentile to $51 at the first percentile).

The average patient received therapy on 14 distinct days during the course of an episode, with a median of 11 therapy days. Total calendar days averaged 55 days with a median of 41 days. PT and OT appear to have shorter courses of therapy, averaging 54 and 50 calendar days, respectively, compared with 71 for SLP. The frequency of weekly visits, measured by the average number of therapy days per week, does not vary much between the disciplines.

CARE-F, on the other hand, had 591 assessments and 518 unique beneficiaries in the nursing facility population. Additionally, 182 assessments and 177 unique beneficiaries are in the day rehab population. The average nursing home patient was 80 years old, whereas the average day rehab patient was 74 years old.

Nursing beneficiaries were composed of a much higher proportion of women (71 percent) than the day rehab population (50 percent). The nursing home population had, on average, a higher proportion of ESRD (2 percent), disabled (11 percent), Medicaid (61 percent), and LTI (80 percent) patients. The day rehab population included 1 percent ESRD, 10 percent disabled, 8 percent Medicaid, and no LTI patients. The nursing population had a much higher composition of Medicaid and LTI patients. Day rehab patients were more likely to be white (90 percent) than nursing facility patients (87 percent), and 10 percent of nursing facility patients were black as compared with 3 percent of day rehab patients.

The average allowed charges for the episode were $3,647 for the nursing population and $4,339 for the day rehab population. Average allowed charges per therapy day were highest for day rehab patients ($187) and lower for nursing facility patients ($102). Episode allowed charges varied 139-fold, from $184 at the first percentile to $25,530 at the 99th percentile for nursing facility patients and 47-fold for day rehab from $391 at the first percentile to $18,293 at the 99th percentile. For nursing facility patients, annual allowed charges were driven mostly by variation in therapy days (207 at the 99th percentile to 2 at the first percentile), rather than by variation in allowed charges per therapy day ($259 at the 99th percentile to $39 at the first percentile). Day rehab patients used fewer therapy days at the extreme (98) and had higher overall charges per therapy day at the 99th percentile ($330).

The average nursing facility patient received 33 distinct days of therapy during the course of an episode with a median value of 20 therapy days. Day rehab patients received an average of 24 therapy days and a median value of 19 therapy days. Total calendar days averaged 60 with a median of 36 days for nursing facility patients and an average of 68 days with a median of 52 days for day rehab patients. The average number of therapy days per week is higher for nursing facility patients (4.02) than for day rehab patients (2.63).

4.6.3 Results: Regressions Predicting Expenditures for CARE-C Community Residents

Payment models for the prediction of therapy resources have low levels of explanatory power (see Tables 7-6– 7-8 in the *DOTPA Payment Alternatives Report*). The PT payment model explains 9 percent of the overall variation in expenditures versus 10 percent for OT and 12 percent for SLP. The payment variables add substantial explanatory power to demographics for all three therapy disciplines. But even with the payment factors, only a small proportion of variation in episode therapy expenditures is explained. The additional factors in the comprehensive model add a significant amount of explanatory power to the payment model, but items such as *pain* and *facility type* are probably not appropriate for a payment model.

Many of the Rasch function subscales predict that higher function is associated with lower therapy expenditures, all other factors equal. Two of the four Rasch function subscales included in the PT payment model—*self-reported participation* and *self-reported mobility*—are statistically significant predictors of episode PT expenditures. Their regression coefficients indicate that a decline from highest ability (score = 100) to lowest ability (score = 0) is associated with a $477 and $248 increase in expenditures, respectively. The Rasch *clinician-reported self-care* and *self-reported life skills* measures are significant predictors of OT episode expenditures. Moving from the highest ability to the lowest ability is associated with a

$533 increase in OT expenditures for *clinician-observed self-care*. However, the opposite direction was found for *self-reported life skills*, with episode expenditures decreasing as patients became more impaired.

Primary diagnoses of *vertigo* and *multiple etiologies (no major)* are negative predictors of PT expenditures, whereas a primary diagnosis of *joint replacement*, *fracture*, and *miscellaneous neurological* are positive predictors of PT expenditures. A secondary diagnosis of *mental health* is a consistent negative predictor of PT expenditures, and *reason for therapy mental function* is a consistent negative predictor for OT expenditures.

Using a *wheelchair/scooter full time* is a strong positive predictor of both PT and OT expenditures, but it is probably not an appropriate payment variable because providers may encourage increased wheelchair use in their patients because of payment incentives. *Using a wheelchair/scooter part time* is not associated with either PT or OT expenditures. *Using a walker* positively predicts PT expenditures, and *using a cane/crutch* positively predicts OT expenditures.

The SLP models are highly problematic because of small sample size—only 125 observations for the payment and comprehensive models. It is difficult to draw any conclusions from the SLP models because of the high likelihood of significant overfitting, multi-collinearity, and outlier influence.

The addition of provider identifiers significantly increases the explanatory power of the models. The episode-based payment model R^2 values increase from 9 to 32 percent, and the comprehensive model values increase from 14 to 33 percent. The sizable increase in R^2 may be due to provider practice patterns, unmeasured case-mix characteristics, or some other unobserved factors that are correlated with the provider. The comprehensive model holds provider type (e.g., private practice) constant; thus, the effect is not due to provider type in the comprehensive model and may be due to individual provider practice.

4.6.4 Patient Panel Size Simulation

To simulate the potential risk of episode-based payment for providers, we evaluated the range of possible payments (the average actual expenditures during the episode) based on the number of PT patients for various provider groups with various patient volumes. There is a very wide range between the minimum and maximum mean expenditures for those patient panel sizes up to 25 patients. Patient panel sizes of 50 or 100 have less variability but still exhibit wide ranges of variation. This level of variation is an indicator of the risk that providers will be exposed to if they have low levels of PT patients in their practice. Patient panel sizes of 500 and 1,000 have the lowest levels of variation and are likely the types of practices that could accept the unsystematic risk of episode-based payments. Note that large patient panel size does NOT reduce the risk of a practice treating a systematically more expensive population such as stroke patients in comparison to those with only sprains/strains. Only risk/case-mix adjustment can mitigate systematic risk.

4.6.5 Results: Regressions Predicting Expenditures for CARE-F Nursing Facility Residents

When the basic payment variables are added to the demographic variables, the explanatory power of the model increases from 4 to 11 percent (adjusted 6 percent; see Table 7-9 in the *DOTPA Payment Alternatives Report*). The basic payment model includes demographic variables, primary diagnosis groups, Rasch clinician-observed mobility and self-care categorical variables, and indicators for diet modification, severe memory impairment, and verbal ability. This limited set of variables is intended to reduce the collinearity that is present in the full payment model because of the overlap between diagnosis, function, structure, and activities groups. In the basic payment model, those beneficiaries younger than 65 years have significantly higher expenditures than the 65–74-year-old reference population. The primary diagnosis groups are insignificant aside from *stroke,* which negatively predicts episode expenditures. The Rasch *self-care* items were positive and significant; the lower the functional score, the higher the predicted episode expenditures. The additional indicators are not significant.

When the additional payment variables are added to the model, the explanatory power increases from 11 to 32 percent (adjusted 21 percent). In this model, none of the Rasch function scales are significant, possibly because of issues of collinearity. Episode expenditures were not affected by prior hospitalization in the past 2 months. Secondary diagnoses of *osteoporosis* or *hypertension* are important predictors of expenditures as was also found in the annual models. *Mental functions* and *motor functions* are important functional predictors of expenditures, and *shoulder/arm/elbow* and *knee* are important positive predictors based on the body structure being treated. *General/no specific body location* and *body structures not reported* are both significant.

The comprehensive model includes additional variables as described in Section 7.2.2 of the *DOTPA Payment Alternatives Report*. The addition of these variables increases the explanatory power of the model to 43 percent (adjusted 30 percent). Of the additional variables added to the model, *having a memory difficulty* is positive and significant, and *having trouble expressing ideas* is a negative and significant predictor of episode expenditures. Unlike the annual model, *prior function on self-care being limited* was not significant.

4.7 Exploring Therapy Case-Mix Groups

Multivariate regression is used in Sections 4, 5, and 7 of the *DOTPA Payment Alternatives Report* to determine which patient characteristics predict outpatient therapy expenditures and how much of the variation they explain. In this section, we use a different statistical methodology—classification and regression tree (CART) analysis—to explore the development of mutually exclusive therapy case-mix groups. CART is a nonparametric technique that does not assume a specific mathematical relationship between expenditures and explanatory variables, and efficiently investigates higher-order interactions (cross-classifications) among the explanatory variables in predicting therapy expenditures.[24] The objective in using CART is to create case-mix groups that have relatively little intra-group payment (resource use) variation and large inter-group variation. We also explore case-mix groups defined from

---

[24] A CART primer is provided in Appendix H of the *DOTPA Payment Alternatives Report*.

clinically and cost-determined primary diagnosis categories, with subclasses based on clinician-observed mobility score ranges.

4.7.1 Methods

The CART analysis was conducted on an episode basis, using the same definition and sample of beneficiaries as used for the regression episode analysis reported in Section 7 of the *DOTPA Payment Alternatives Report* and summarized previously in ***Section 4.6***. Discipline-specific CART analyses were performed for the PT, OT, and SLP patients using the comprehensive and payment models.

For PT patients only, a primary diagnosis model was also estimated using CART. A primary diagnosis model was used because diagnosis is usually the first stage in case-mix classification of individuals. The grouping of the primary diagnosis categories is on statistical grounds only and might require adjustment for clinical face validity. Therefore, without use of CART, we explored creating payment groups by collapsing the 21 primary diagnosis groups (described in ***Section 3***) into 12 diagnosis-aggregated groups based on clinical and cost similarity. Then, we examined subgroups within the 12 diagnosis groups defined by clinician-observed mobility.

4.7.2 Results: CART Analyses of Case-Mix Groups

The result of the PT CART analysis including the widest range of variables (comprehensive model) shows *patient use of a walker* and *full-time wheelchair use* to be among the most important variables defining case-mix groups. This indicates that beneficiary mobility and perhaps frailty is important in defining case-mix. For payment, it may be preferable to measure mobility using clinician assessment or patient self-report and not on patient utilization of medical services or devices. Another utilization-related variable—*timing of surgery*—is also important in defining case-mix. Nevertheless, even the comprehensive model of PT case-mix achieves a maximum cross-validated R^2 of only 3.2 percent.[25] The conclusion is that the vast majority of variation in PT episode expenditures is not explained by the available clinical, patient, and case-mix factors.

When a smaller set of payment case-mix variables are employed, *timing of surgery* followed by *clinician-observed mobility* and *self-reported participation* are most important. This is consistent with the stepwise regression analysis presented in Section 5 of the *DOTPA Payment Alternatives Report* and summarized in this report within ***Section 4.4***. The maximum cross-validated R^2 using the payment variables is 1.7 percent, which is about half of what is attained with the larger comprehensive set of variables.

---

[25] The cross-validated R^2 of 3.2 percent for the CART PT comprehensive model is much lower than the adjusted R^2 of 8.7 percent for the ordinary least squares (OLS) PT payment presented in Table 7-6 of the *DOTPA Payment Alternatives Report*. CART cross-validated R^2s are lower than the OLS adjusted R^2s in part because of how CART creates interacted variables, the accompanying validation process, and the minimum sample size needed to form the terminal nodes. A sample size larger than the one obtained for DOTPA could have permitted the creation of more homogeneous groups and thereby produced a better fitting model.

Primary diagnosis is a key variable in many existing Medicare case-mix systems and therefore is chosen as an initial case-mix classification variable. However, the unguided CART classifications did not statistically choose to begin groupings with diagnosis; therefore, we examined PT case-mix groups in which CART was constrained to begin with primary diagnosis. This CART analysis of the 21 mutually exclusive primary diagnosis groups yielded 8 aggregated diagnosis groups and achieved a cross-validated R^2 of 2.9 percent, which approaches the cross-validated R^2 achieved from the comprehensive model.

In the OT CART analyses, *full-time wheelchair use* was the most important variable in the comprehensive model, and *self-reported everyday activities* was the most important among the variables considered suitable for payment. In the SLP CART analysis, *clinician-observed problem-solving skills* was the most important variable in both the comprehensive and payment models.

4.7.3 Results: Primary Diagnosis and Clinician-Observed Mobility Physical Therapy Case-Mix Groups

We also defined 12 aggregated PT primary diagnosis groups from the initial 21 groups using clinical and cost criteria (without using CART) and cross-classified them with pre-defined clinician-observed mobility categories. We split 5 of the 12 diagnostic groups into 2 groups using Rasch mobility estimates of 0–70 and 70–100 on a scale of 0 (lowest mobility) to 100 (highest mobility). This process resulted in 17 final case-mix groups: 7 based on diagnosis alone and 10 based on 5 diagnosis groups split into higher and lower mobility subgroups. These 17 groups explained 4.2 percent of the overall variation in PT episode expenditures.

4.7.4 Implications

Like many of the analyses based on the CARE-C sample of beneficiaries, the small sample size limited what could be learned from the use of CART and methods using primary diagnosis to initially classify beneficiaries into payment groups. The result of having few terminal nodes (case-mix groups) for the comprehensive and payment models is probably partly due to the relatively small overall sample sizes for each of the three disciplines, because CART generally requires larger sample sizes than more familiar multivariate techniques such as ordinary or generalized least squares. An unconstrained CART did not use primary diagnosis as the initial classification variable. In future work, a constrained version of CART could specify primary diagnosis as the initial classification variable and then include other relevant variables.

As an alternative to CART, we defined 12 aggregated PT primary diagnosis groups using clinical and cost criteria and cross-classified them with pre-defined clinician-observed mobility categories. This process assigned a greater role to clinical judgment, payment policy considerations, simplicity, and face validity than the CART analyses, and it used only two key variables—*primary diagnosis* and *clinician-observed mobility*—from among the large number of potential case-mix variables. This method could be expanded with more variables and more splits, but, as is true of all the case-mix analyses, it would be limited by available sample sizes.

4.8 Concluding Observations

Medicare outpatient therapy is characterized by a high degree of variability in expenditures among users on an annual or episode basis. A major goal of this report is to determine whether a large range of measurable factors—in particular, patient functional status—explain that variation. We found that patient functional status and other variables do not explain more than a modest amount of variation in outpatient therapy expenditures. It appears that substantial variation in outpatient therapy expenditures is unrelated to patient "need" for therapy as defined by these variables. It is possible that clinical-need factors that we did not measure could explain more of the variation in expenditures, but we measured a wide range of patient-reported, clinician-observed, and administrative data and produced a number of diagnostic, functional, clinical, and utilization variables. Further refinement of case-mix measurement would benefit from larger sample sizes for analysis, especially for OT, SLP, and nursing facility patients. The case-mix measurement could also benefit from enhancements to the assessment items to resolve issues such as ceiling effects and nonspecific but statistically significant results in the nursing facility models.

Although the measured case-mix factors do not explain much therapy expenditure variation, case-mix adjustment of Medicare therapy caps or episode-based payments may nevertheless be desirable. We identify a number of characteristics that do predict some degree of expenditure variation, including diagnosis, functional status, and utilization. Case-mix groups defined by these factors predict expenditure differences at least on the order of three to one, which seems to indicate some group-level differences in expenditures. Case-mix adjustment of therapy expenditure caps or episode-based payments could be approached using only administrative data or by also incorporating patient self-report and clinician-assessment data. The former does not include systematic information on patient functional status and other characteristics, whereas the latter requires a potentially expensive and burdensome primary data collection system. Even with case-mix adjustment, however, the large amount of unexplained variation in therapy expenditures will require additional strategies in reimbursement. These strategies could include renewable fixed-length episodes, mixed episode/FFS payment approaches, and outlier policies.

In addition to case-mix adjustment, a possible refinement of Medicare's annual therapy expenditure caps is to impose them on a discipline-specific or a combined-disciplines basis, rather than the current combined cap for PT and SLP and separate cap for OT. A budget-neutral, equal discipline-specific cap would be slightly lower than the current cap, whereas a combined-disciplines cap would be higher. Discipline-specific caps are most generous for beneficiaries needing a large amount of multiple disciplines of therapy, whereas a combined cap is most generous for beneficiaries needing only one discipline. Our budget-neutral analysis indicated that a combined cap would apply to a smaller portion of beneficiaries, but affected beneficiaries would exceed the cap by a larger amount, on average. A combined cap could reduce the burden of CMS or contractor manual medical necessity review because of a smaller number of beneficiaries exceeding the cap.

[This page intentionally left blank.]

5. LESSONS LEARNED

This section summarizes lessons learned from the Developing Outpatient Therapy Payment Alternatives (DOTPA) project. ***Section 5.1*** discusses lessons for Medicare outpatient therapy payment, including adjusting the therapy cap for beneficiary needs, cap policy for the three therapy disciplines, episode-based payment, and blended fee-for-service (FFS)/episode-based payment. ***Section 5.2*** discusses lessons learned regarding the Continuity Assessment Record and Evaluation (CARE) instrument items. ***Section 5.3*** discusses lessons learned for collecting data to support therapy case-mix measurement such as the collection of patient self-report items in addition to clinician-assessment data; this section also includes commentary on data collection procedures, missing data, and sample sizes necessary for case-mix analysis.

5.1 Lessons Learned for Alternative Approaches to Therapy Payment

In this report, we considered three alternative approaches to the current Medicare payment system for outpatient therapy: (1) refining the cap on annual expenditures, (2) episode-based payment, and (3) mixed (blended) FFS and episode-based payment. Following is a discussion of the lessons learned for each approach.

5.1.1 Refinement of the Annual Therapy Expenditure Cap

Adjusting the Annual Cap for Beneficiary Need for Therapy

We examined adjusting the annual expenditure cap based on beneficiary need for therapy. We tested a wide range of administrative, clinician assessment, and patient self-report data as predictors of beneficiary annual therapy expenditures (among beneficiaries with some use of outpatient therapy). These variables included demographics, diagnoses (from claims and assessments), motor and cognitive functional status (from clinician assessments and patient self-report), hospitalizations by diagnosis-related groups, inpatient utilization of therapy, reasons for therapy (body structure, body function, activity limitations), pain (patient self-report), and surgical history. Separate models were estimated for physical therapy (PT), occupational therapy (OT), and speech-language pathology (SLP) in the CARE-C setting, whereas a combined model was estimated for the three disciplines in the CARE-F nursing facility setting. Both ordinary least squares models predicting mean expenditures and quantile regressions predicting the expenditure cap were estimated. We estimated comprehensive models in an attempt to isolate the "true" marginal effect of predictors on expenditures, holding as many other factors constant as possible. These models included *payment models* that eliminated variables from the comprehensive model thought to be unsuitable for payment; *basic models* that included only the key predictors diagnosis and functional status; and, to address redundancy among the predictors, stepwise regressions that added variables to the model one by one based on their incremental predictive power.

One conclusion from these models is that systematic differences in annual therapy expenditures by some of these characteristics were observed. The differences in average (or quantile) therapy expenditures associated with selected beneficiary characteristics could be used to adjust the annual therapy expenditure cap. Some of the factors predicting therapy expenditures are available from existing administrative data, whereas clinician assessment and patient self-report data would have to be obtained from a costly new and ongoing data collection effort from outpatient therapy providers and/or patients.

A second conclusion is that the amount of variation in individual beneficiary annual therapy expenditures explained by the measured factors is small. This means that either (1) therapy expenditures are closely related to need for therapy, but available data and models are not able to estimate individual beneficiary therapy needs effectively; or (2) that individual beneficiary therapy expenditures are not closely related to the amount of therapy needed.[26] The weak ability to accurately predict the current dollar volume of therapy utilization in relation to characteristics we measured implies that even with a risk-adjusted cap, there will continue to be substantial demand for exceptions to the cap. This suggests that it will continue to be difficult for Medicare to impose a "hard" cap (without exceptions) on therapy expenditures, but risk adjustment will be able to modestly improve the targeting of therapy manual review. Further work to improve the measurement scales developed in this project, such as lessening or eliminating ceiling effects, could improve the ability to explain expenditures, although large gains do not seem likely.

Current Policy versus Discipline-Specific versus a Combined-Discipline Cap

We also examined how the therapy cap is imposed by therapy discipline. Current policy imposes the cap for PT/SLP combined and separately (and equally) for OT. We considered two alternative policies: (1) separate and equal caps for each of PT, OT, and SLP; and (2) a single combined cap for PT, OT, and SLP.

Using existing claims data, we simulated versions of the two alternatives that are budget neutral with respect to current policy, assuming no behavioral response on the part of beneficiaries or providers to cap changes (see ***Section 4.3***). Under the two alternatives, subgroups of beneficiaries fare differently. Equal, budget-neutral discipline-specific caps are each slightly lower than either of the two current policy caps, and a slightly higher percentage of beneficiaries are above at least one cap. Discipline-specific caps are most favorable to the minority of beneficiaries using multiple (especially all three) therapy disciplines because these beneficiaries can "max out" each of the three caps separately and achieve higher total therapy spending under the cap than with the other two policy alternatives. A single combined cap is higher than the current policy caps, and a significantly lower percentage of beneficiaries would exceed the single combined cap. A combined cap is most favorable to beneficiaries using only a single therapy discipline. Discipline-specific caps match more closely with the actual pattern of therapy expenditures by users of one discipline, two disciplines, or all three disciplines than current policy or a combined cap. A combined cap would target for review a smaller number of high outpatient therapy users across all disciplines.

On balance, members of the technical expert panel (TEP) convened in December 2013 favored the idea of moving to three equal, discipline-specific caps. Of the 11 TEP members who

---

[26] The hypothesis that expenditures may not be closely related to the amount of therapy actually required suggests possible over-utilization of service. Evaluation of the DOTPA discharge assessments may shed some light on the adequacy of the therapy services actually provided but would not necessarily identify excessive use. Such an issue should be considered for future research toward an alternative payment system.

preferred a particular cap policy, 8 favored three equal, discipline-specific caps, whereas 3 favored a single combined cap. Many TEP members opposed therapy caps in any form; several TEP members commented that they saw no rationale for the current policy of a combined PT/SLP cap and a separate OT cap. Those favoring a combined cap cited the substitutability among PT and OT services.

5.1.2 Creation of an Episode-Based Payment

Distribution of Outpatient Therapy Episodes and Unit of Payment

We examined the pattern of Medicare episodes of outpatient therapy treatment by constructing variable-length episodes using 100 percent claims data. We found a high degree of variability of Medicare episode-based payments, driven primarily by the length of the episodes. Episodes ranged in length from 1 day to longer than 1 year. In payment, the variability of episode length will need to be dealt with, most likely by renewable fixed-length episodes or possibly by an outlier policy. The unit of episode-based payment for outpatient therapy could be, for example, renewable 30-, 60-, or 90-day episodes.

We analyzed 30-, 60-, and 90-day initial fixed-length episodes for physical therapy beginning in 2010 for community residents (i.e., the CARE-C population). The 30-day fixed episode length truncates many initial therapy courses of treatment and accounts for only 40 percent of all Medicare payment for PT beginning in 2010. Fifty-five and 63 percent of payments for all PT episodes beginning in 2010 occur within the 60- and 90-day fixed windows of the initial episodes, respectively. These results suggest that a fixed episode length longer than 30 days—perhaps 60 or 90 days—would be most appropriate for outpatient therapy payment.

Case-Mix Adjustment of Episode-Based Payments

We examined predicting episode expenditures using clinician assessment and patient report information, combined with data from Medicare administrative enrollment and claims files. Predictors tested included demographics, diagnoses, functional status, reasons for therapy (body structure, body function, and activity limitations), patient self-report of pain, and surgical history. Separate models were estimated for PT, OT, and SLP. As found for annual expenditures, we also found that some factors predicted systematic average differences in therapy episode expenditures. But, as a group, the predictors accounted for only a small portion of the overall variation in expenditures.

The implication for episode-based payment is that some case-mix adjustment variables are available—for example, certain diagnoses and functional status—that predict meaningful and statistically significant differences in expenditures in large groups of beneficiaries. But the available case-mix variables are weak predictors of therapy expenditures in individual episodes. Therefore, small outpatient therapy practices—in which much of Medicare therapy takes place—would be subject to considerable financial risk under an episode-based payment system, even if it were case-mix adjusted.

5.1.3 Creation of a Mixed (Blended) Episode and Fee-for-Service Payment

As discussed previously, feasible case-mix adjustment for outpatient therapy with currently available variables is “weak” in the sense that only a small portion of expenditure variation is explained. Weak case-mix adjustment creates financial risk for providers and allows greater scope for providers to enhance their profitability by selectively enrolling healthier patients. One payment approach to mitigate risk and selection incentives is a blend of both FFS and episode-based payment. In mixed payment, therapy providers are paid partly on an episode lump-sum basis and partly on a fee-for-service basis for each additional therapy service provided. For example, the episode lump-sum payment might be $150 and payment for each service could be 70 percent of the current fee schedule amount. Mixed payment partly retains episode-based payment incentives for greater efficiency while using FFS payments to reduce provider risk and selection incentives.

We simulated current FFS, full episode, and mixed FFS/episode-based payment for initial fixed-length PT episodes of care among community residents beginning in 2010. We showed how two aspects of mixed fixed-length episode-based payment could reduce provider risk and redistributions from current FFS payment: (1) a higher proportion of FFS payment (e.g., 70 percent rather than 50 percent), and (2) a shorter fixed episode length (e.g., 60 days rather than 90 days). We conclude that the Centers for Medicare & Medicaid Services could consider mixed payment for outpatient therapy as a means of addressing the weak case-mix adjustment that is possible.

5.2 Lessons Learned about the CARE Standardized Items

Several types of lessons were learned from the use of the standardized items to collect data on the Part B therapy populations. Several of the lessons focused on design of the data collection instrument, whereas other areas focused on ways to improve the items included in the data collection instrument.

5.2.1 The CARE Instrument

Data Reduction

The test instrument was useful for identifying the best way to collect data about a particular characteristic, but the final set of items to include in an actual data collection instrument should be limited to one item per attribute. For example, the DOTPA tool had several items measuring presence and severity of pain, including but not limited to the primary diagnosis, secondary diagnosis, reason for therapy, and limitations of certain activities because of pain. If this type of instrument moves forward, item reduction should establish only the most valid and reliable measure of such concepts and should be consistent with current clinical practice and workflow. An exception is the function scales (***Section 3***), which require several items to estimate functional ability scores.

In discipline-specific CARE-C analyses, only a subset of available Rasch function scales was used. The function scale subsets for discipline-specific analyses were selected for analysis based on relevance to the scope of practice and training of the discipline, typical functional limitations associated with patients being treated by the discipline, and the amount of non-missing data. In the future, either discipline-specific CARE-C versions could be created such that each version only includes items relevant to the specific discipline, or so that non-relevant items could be skipped using an electronic-based data collection instrument using the characteristics of the therapist. For instance, the Speech and Functional Voice scales are primarily relevant to SLP practice and were used only in SLP analyses; these scales could be excluded from the PT and OT CARE-C instrument versions.

Skip Patterns

Skip patterns should be revisited in the future. The skip pattern in the CARE-F assessment, Section III, item B2 (how many surgeries has the beneficiary had?) has two skip options. This pattern should be corrected to just one skip option. Conversely, the skip pattern on the mobility questions may need to be augmented to include more than one skip option (for CARE-C as well as CARE-F). The mobility skip question also needs to include bed mobility/transfers to ensure that beneficiaries skipping the mobility scale do not have limitations on any of the mobility items. These skip patterns should be revised, as recommended, before implementation. In addition, within the CARE-C instrument, clinician-reported self-care and mobility items should have a skip pattern included to prevent ceiling effects due to beneficiary limitations that are irrelevant for the particular subscale. This recommended skip pattern methodology is similar to what is already in place for the patient self-report items.

5.2.2 Item Specification

Several issues emerged related to item specification. Items could be condensed or expanded, depending on the part of the tool.

Function Scales

Many of the function scales in the CARE-C instrument suffered from ceiling effects (large proportions of therapy users clustered at the highest or unimpaired scale values) in the community population. In future work, emphasis should be put on developing scales less subject to ceiling effects in this population by including higher-difficulty items. For example, more Instrumental Activities of Daily Living subscale items could be included, or walking/wheelchair distances that more accurately portray an independent trip to the store (e.g., parking lot distance, obstacle maneuvering, standing and walking for a period of time), rather than 500 feet in a straight line. However, the therapist burden should be carefully considered in the development of any new items.

Finally, in both the CARE-C and CARE-F clinician-reported Mobility subscales, the data collected on the distance items did not allow for an in-depth assessment of how well the items agree between expected and observed responses. Because therapists responded to only one item, it was difficult to determine whether someone who needed partial assistance on the furthest distance could be considered independent on the shortest distance. These items need to be revisited in future work with different data collection rules.

Response Options for Function Items

The CARE DOTPA data suffered from a large amount of missing data on the clinician-assessment items; this problem was addressed—primarily by using only scales with sufficient non-missing data in analyses and/or including "missing" as a variable response category in analyses—but only after data collection during the analysis phase. Creating discipline-specific CARE-C versions that include only discipline-relevant items, as suggested previously, would help limit the data requested to what is feasible and relevant for the respondent to provide. Additionally, in the future, the reasons for nonresponse should be more carefully collected, and collapsing of multiple responses into a single rating scale category/letter code should be avoided. On CARE-C assessments, the Not Assessed response option under the clinician-observed function scales currently includes both *not clinically relevant* and *outside scope of knowledge/skills*. This response option could be revised such that *not clinically relevant* and *outside scope of knowledge/skills* are separate categories as follows: (1) patient is functionally limited, within scope of practice, but not being addressed; or (2) outside scope of practice. Additional CARE-C Not Assessed response categories could include *medical/safety concerns, patient refused*, and *not a treatment goal*. Similarly for CARE-F assessments, the Not Assessed response category for clinician-observed function items could be revised to include (1) limited, within scope of practice, but not being addressed; (2) outside scope of practice; (3) medical or safety concerns; or (4) environmental constraints. Additional categories for *patient refused* and *not a treatment goal* could also be considered for inclusion in CARE-F clinician-observed function items.

Mode of Observation

In both the CARE-C and CARE-F clinician-reported subscales, further detail on the mode of observation is needed. In many cases, the therapist observed an approximation of the task (e.g., hands over head for *dress upper body*) and coded a dependence level on this task-relevant behavior. If the collection of the various observation modes is deemed important and not overly burdensome to collect, further research is needed to collect the details on how those task-relevant activities differ from the task actually performed.

Diagnosis List

The diagnosis list represented medical diagnosis but omitted many rehabilitation diagnoses (e.g., lower back pain) that could explain the cause of the functional limitation. Future development of this tool should expand the list of diagnoses that could explain the underlying etiology or cause of the functional limitation. Relevant medical diagnoses such as spinal cord injury, adhesive capsulitis, surgical capsular release, impingement syndrome, and amyotrophic lateral sclerosis should be added.[27]

---

[27] The TEP recommended classifying patients into diagnosis groups based on the underlying etiologic diagnosis (stroke, hip fracture) to understand the cause of the current impairments or rehabilitation problems, such as pain or paralysis or aphasia. Updating the diagnosis list would allow capturing the underlying etiologic diagnosis more consistently, when this information is known.

Primary and Secondary Diagnosis

This section was particularly difficult to analyze. Considerable reassignments needed to be done to combine these variables consistently. The list of primary diagnoses should be revised so that the reassignments of primary diagnoses that we did are not necessary. Future work should allow the primary and secondary diagnosis concepts to be separately measured, with the primary diagnosis reflecting the underlying etiological reason for treatment and the secondary diagnosis reflecting coexisting comorbidities or complications. Additionally, diagnoses that are unlikely to be etiological in nature for the specific discipline treating the patient, such as depression and hypertension in PT, should be removed from the primary diagnosis list and listed as a secondary diagnosis to reflect their status as comorbidities.[28] To accurately classify the diagnoses for beneficiaries presenting with impairments and/or symptoms for which an underlying etiological diagnosis cannot be identified, consider adding Not Otherwise Specified to impairment/symptom primary diagnosis labels so they are selected only when the underlying etiological diagnosis is unknown (e.g., pain – Not Otherwise Specified, gait or balance disorder – Not Otherwise Specified, generalized weakness – Not Otherwise Specified).

Rehabilitation/symptom/impairment diagnoses also contain information and should also be collected, but they should be clearly distinguished from underlying medical etiological diagnoses (i.e., there should be separate categories of response for the two types of diagnoses). For example, *low back pain* could be reported as the rehabilitative diagnosis, in association with osteoarthritis, sprain or strain, or some other “unknown” malady as the medical etiological diagnosis.

While the primary and secondary diagnosis information is collected differently in different settings, the results clearly showed the importance of separating primary and secondary diagnosis information to understand the causes of the functional limitation separate from the complications affecting the therapy outcomes. Further consideration needs to be given regarding whether therapists can accurately report all diagnoses, especially comorbidities not directly relevant to the problem under treatment, or whether those diagnoses should be obtained from other sources, such as Medicare claims.

Although the CARE instructions stated providers should select a single primary diagnosis, several providers reported more than one primary diagnosis for each CARE assessment completed. This required complex reassignment procedures to various “multiple diagnoses” groups. In the future, the rule that only a single primary diagnosis be reported should be enforced. A single principal diagnosis greatly simplifies case-mix classification, and the reporting therapist is in the best position to judge what the principal diagnosis relevant to the therapy plan of care is. If certain combinations of diagnoses are highly salient and frequent for therapy treatment, they could possibly be explicitly listed as primary diagnosis options. However, allowing multiple primary diagnoses creates too many case-mix classification options to be feasible; therefore, possible combinations of primary diagnoses need to be limited. If

---

[28] Inconsistent or inappropriate response combinations for primary and secondary diagnoses could be addressed through an electronic CARE assessment tool. Such combinations would need to be defined based on additional research and consultation with clinical staff.

electronic data collection forms are implemented, checks should be incorporated to prevent the provider from moving to the next item if he or she selects more than one primary diagnosis.

Reason for Therapy

Although conceptually important in understanding therapy needs, we found that the CARE *reason for therapy* variable did not perform well in many of our models, probably in part because of redundancy with other variables. This variable is another item that needs further refinement. In particular, the participation items were not consistently reported by therapists, some were too general or vague to be used for payment, and some were redundant with the self-reported participation items. It was difficult to analyze multiple reasons for therapy. Consideration should be given to collecting the primary or single most important reason for therapy. Additional reasons for therapy items could potentially be collected as secondary reasons for therapy.

Vague or non-specific categories such as *Other* or *General/No Specific Body Location* were frequently chosen as responses, and some of them were occasionally important in explaining therapy expenditures. These categories could potentially be indicators of general frailty. Consideration needs to be given to interpreting these categories and potentially refining them to provide more specific information. Feedback from the TEP supported the suggestion that these non-specific responses could be an indication of general patient frailty.

Mobility Devices and Aids

For wheelchair users, it would be helpful to have information regarding the type of wheelchair used (i.e., manual versus motorized). Use of a motorized scooter can indicate a higher degree of impairment (e.g., less upper-body strength to move the wheelchair).

Cognition Measurement

Beneficiary cognitive status is widely believed to be an important factor that may influence therapy utilization, as evidenced by TEP member discussions. Impaired cognitive status is clinically expected to increase resource needs; for example, treating and educating patients with cognitive impairment may take longer compared with patients who have intact cognition. On the other hand, the subgroup of patients with very severe cognitive impairment may have lower therapy utilization given their limited potential for improvement or learning; these patients may instead primarily require caregiver education.

Although expected to be an important predictor, neither self-reported nor clinician-observed cognitive impairment, as measured by the Rasch function scales, significantly predicted higher therapy expenditures in this project. Self-report of functional status, particularly cognitive function, can be unreliable in patients with cognitive problems, given that their ability to accurately self-assess their functional status may be impaired. For this reason, the CARE instrument instructed clinicians to use a proxy reporter if the patient was unable to provide accurate information about his or her health. Nevertheless, contrary to expectation, higher ability in the self-reported cognition/communication scale (i.e., Life Skills scale) predicted significantly higher therapy expenditures in the CARE-C OT model. Studies on the Activity Measure for Post-Acute Care have also shown lower subject-proxy reliability of the Applied Cognitive scale, compared with the Basic Mobility and Daily Activities scales. Therefore, we recommend that self-reported cognitive function not be used for predicting therapy utilization.

The clinician-observed Cognition scales (problem solving, memory, and attention) could not be used in the CARE-F nursing models because these items were completed for only beneficiaries being treated for a cognitive problem rather than for beneficiaries who had a cognitive problem. Thus, cognitive scores of beneficiaries for whom the cognitive scales were skipped could not be recoded to maximal ability. If it is desirable to use clinician-observed cognitive status in future payment analyses, the CARE-F assessment should be revised such that the clinician-observed cognitive-scale items are completed for all beneficiaries. Item III.IV.B3 on the CARE-F nursing facility assessment[29] would be the more appropriate gateway question for the clinician-observed Cognition scales with some minor text modification.

Other CARE instrument items that indicate possible patient cognitive problems were also included in payment analyses. These items include selected primary and secondary diagnosis groups, select primary reason for therapy items, screening questions for cognitive problems in the CARE-C PT and OT samples, and a global rating of cognitive function in the CARE-F nursing facility sample. The cognitive screening questions did not measure the severity of the cognitive impairment; instead, they measured only the impairment's presence or absence. These items were used in CARE-C PT and OT analyses where the clinician-observed Cognition scales could not be used. The global rating of cognitive function item (mildly impaired, moderately impaired, or severely impaired) was analyzed in the CARE-F sample in which the clinician-observed Cognition scales could not be used; however, it was often missing or not assessed, especially in the CARE-C sample, precluding examination of the severity of the impairment in the community PT and OT samples. Although some primary reason for therapy and cognitive status variables were significant predictors in some models, no consistent association was found between any of the CARE instrument cognitive variables and therapy expenditures.

It is recommended that alternative approaches for assessment of cognitive function in the context of predicting Part B therapy expenditures be explored in the future. Additionally, the relationship of the Brief Interview for Mental Status and Confusion Assessment Method, included on the CARE-F assessment, with therapy expenditures should be explored.

5.3 Lessons Learned about Collecting Data to Support Therapy Payment and Case-Mix Adjustment

A key feature of the DOTPA project was the collection of data on therapy patients either directly or indirectly via a proxy respondent. Therapy payment reforms and case-mix and outcomes measurement will likely require collection of similar data. Here, we offer observations from our experience and analysis on the usefulness of patient self-report data in addition to clinician assessment data, procedures for collecting the data, and issues with missing data among

---

[29] This item is "Does the patient have any difficulty with memory, attention, problem solving, planning, organizing, or judgment?"

the collected data. We also discuss what we learned about sample sizes necessary for developing therapy case-mix adjustment models.

5.3.1 Collecting Patient Self-Report Data in Addition to Clinician-Observed Data

The DOTPA project collected patient self-report data and clinician –observed data from therapy sites of care. Collecting the patient self-report data incurs costs and imposes burdens on patients. An important question for future data collection is whether the usefulness of the patient self-report data justifies the cost and burden of collecting the information.

Twelve of the 14 TEP members responded that they prefer to include both clinician-observed and self-reported items in a case-mix model to predict therapy expenditures. Overall, the TEP thought that both clinician-observed and self-reported items were complementary and should be combined to obtain a more complete picture of the patient's functional ability. Several members pointed out that the self-reported items would not be appropriate for patients with cognitive impairments.

Analyses from the *DOTPA Measurement Report* (Kline et al., 2014) found some item similarity between the patient self-report and clinician-observed items, but low correlations and low intra-class correlations led to the recommendation of including both scales to capture potentially different aspects of function. TEP comments and the recommendations from the *DOTPA Measurement Report* were also validated by empirical analysis in Section 5.4.1 of the *DOTPA Payment Alternatives Report* (Amico et al., 2014b). The explanatory power of the model including both the clinician-observed and self-reported measures was greater than either the clinician-observed mobility scale or the self-report measures alone. This supports the assertion that different information is being collected in both types of measure. Although there is value to collecting both self-report and clinician-observed measures, future work should consider the added cost of collecting these measures in relation to the benefit of added explanatory power.

5.3.2 Data Collection Procedures

The DOTPA data collection began with recruitment of participating Medicare outpatient therapy providers. This effort was originally scheduled for several months before the collection of any assessments but proved more difficult than was anticipated. Many providers who declined to participate cited the length of the assessment instrument and a lack of sufficient time and resources as their reasons for doing so.[30] Difficulties in recruitment led the DOTPA team to extend the enrollment period into the data collection phase and to revise enrollment targets downward. Future data collection efforts should consider a shorter survey instrument to minimize respondent burden or at least an electronic data collection instrument with appropriate skip patterns to lower burden.

An individual provider's participation in DOTPA began with a Web-based training session instructing them in both the data collection protocol and the clinical content of the assessment instrument. Although Webinars proved much less costly than in-person training

---

[30] Some participating provider sites later indicated that skip patterns on the instrument made its completion easier than they had initially anticipated.

sessions, they may have been less effective. As with most Webinars, it can be difficult to monitor the engagement of the participants. Often, completed assessments were returned to the DOTPA team with errors that were discussed at length during training. Future efforts should consider more interactive approaches to training from a distance that maximize participant engagement, as well as whether in-person training is worth the additional cost.

Assessments were collected on paper forms and returned by mail every 2 weeks. As such, an assessment was typically 2–3 weeks old before the DOTPA team could review the completeness of the responses. Furthermore, because some assessment data were captured with a paper form, initial review was limited to the structure of the responses (e.g., date of birth being recorded in the form mm/dd/yyyy), and immediate validation of the content with other data sources such as claims or Medicare enrollment files was not feasible. Many of the errors discovered during this review were mistakes in protocol (e.g., incorrectly following a skip pattern or leaving a required question blank). Any mistakes were immediately communicated to the provider site coordinator. Often, corrections were obtained from these providers, although some could not locate the missing or correct information (usually citing a lack of time).

The DOTPA team then transcribed completed assessments into electronic databases. This proved a much more burdensome and costly task than was originally anticipated. Transcription was partially automated by optical character recognition software as originally planned, but output was at times unreliable and required manual validation. Furthermore, transcription of alphanumeric and free-text responses (e.g., Medicare Health Insurance Claim Number, date of birth, National Provider Identifier) could not be automated and required manual data entry. This time-consuming effort delayed completion of the analytic data file and detailed validation of the data contained within the assessments.

Future efforts should use electronic methods of data collection that could both automatically generate analytic databases and validate responses in real time. These methods would effectively eliminate burdensome transcription and review efforts and prevent providers from submitting protocol mistakes entirely. The system could also automate skip patterns in the assessment instrument. Skip patterns alleviated some of the anticipated participant burden. Although automation would not further reduce this burden, it could present a less challenging first impression during recruitment and boost the effectiveness of that effort. Such a system was tested with a small subset of DOTPA participants during data collection, and the effort was largely successful.

5.3.3 Missing Item Responses

In addition to the overall participant sample sizes, many of the individual items on the CARE assessment were not answered, thus making it impossible to construct the patient function scales. In these cases, we dropped many cases from the analysis to run the regression models. Some of the nonreporting may have been because responding therapists did not see certain CARE questions as relevant to their treatment of the patient or did not feel qualified to provide the requested information. For example, these reasons may have been behind the low response rate to the CARE self-care items by physical therapy patients. As discussed in ***Section 5.2.2***, there was also confusion about the Not Assessed items that had only one response option but could have represented three different reasons for clinicians’ inability to assess. Restructuring the CARE questionnaire so that it asks for reporting of only data that are feasible and relevant for the respondent to provide, as well as including multiple response options that specify the reason data were not provided, could lessen the degree of missing or unusable data in future data collection efforts. As implemented for DOTPA, the revised instrument would need to be tested on a small set of providers and beneficiaries to ensure the changes were as intended.

5.3.4 Sample Sizes

Because of the issues noted previously, the project did not have the intended sample size available for analysis. OT had only about 500 CARE-C episodes or beneficiary/years for analysis, and SLP had fewer than 200. Sample sizes for the combined-disciplines nursing facility analysis were 500–600 episodes or beneficiary/years. These very small sample sizes meant that many of our OT and SLP community case-mix regression models and our nursing facility models overfit the data and did not achieve valid and reproducible results. We had far too many explanatory variables in the OT, SLP, and nursing facility models relative to sample size, so we could not draw many solid conclusions about the independent effect of those variables on therapy expenditures. The classification and regression tree (CART) analysis was even more limited by small sample sizes. Mutually exclusive case-mix group sizes, and therefore, statistical precision in measuring expenditure differences, rapidly diminish in the CART analysis because case-mix groups are iteratively split by additional variables.

The PT sample size of more than 4,000 CARE-C episodes or beneficiary/years was much better. Nevertheless, the most expensive cases—the ones most likely to exceed the therapy caps—tend to be rare and were poorly represented in the PT sample. The small sample sizes of expensive cases limit the ability to identify and develop case-mix categories to adjust for them. Future work should consider oversampling the rare, expensive cases, such as paralysis, amputation, burns, severe head injuries, and severe neurological disorders (e.g., quadriplegic cerebral palsy, amyotrophic lateral sclerosis) to identify and develop case-mix categories to adjust for these costly groups. Future CART work to create mutually exclusive case-mix groups will need to have much larger sample sizes to achieve validity and reproducibility.

Without attempting a formal statistical power analysis, we opine that a minimum of 5,000 episodes or beneficiary/years is necessary to develop a valid basic case-mix model. To develop a refined model, random samples of at least tens of thousands of cases, or smaller random samples with an oversample of high-need conditions or beneficiaries, are required.

6. SUGGESTIONS FOR FUTURE RESEARCH

In this section, we offer suggestions for further research on outpatient therapy payment and case-mix adjustment. We discuss future research ideas related to function measures and scale development (***Section 6.1***), conditions and diagnoses (***Section 6.2***), other factors affecting therapy costs (***Section 6.3***), and model specification and estimation (***Section 6.4***). Additional suggestions related to issues outside the scope of the Developing Outpatient Therapy Payment Alternatives (DOTPA) project are provided in ***Section 6.5***.

6.1 Function Measures and Scale Development

Developing measures of beneficiary function and need for therapy was a primary purpose of the DOTPA project. The project focused on developing Rasch scales of function, but other approaches are possible. We list and briefly discuss the following research ideas.

1) *Compare the performance of additive function scales with the performance of Rasch function scales.*

Additive scales simply add up a respondent's score on individual scale item. They are less sophisticated than Rasch scales, but they are also simpler to compute and comprehend. The performance of additive versus Rasch scales in predicting therapy expenditures could be compared to help judge whether the predictive gains from Rasch scales are worth their added complexity.

2) *Examine the performance of individual function items in explaining therapy expenditures, and create an expenditure-weighted scale.*

In the DOTPA project, we examined how multi-variable scales of patient functioning predicted therapy expenditures. The individual variables were combined to form scales based on psychometric (e.g., Rasch) analysis. Multi-variable scales are thought to measure underlying constructs more precisely and reliably. Less random measurement error should be associated with better prediction of therapy expenditures.

However, because the ultimate goal is to predict therapy expenditures, an alternative would be to regress therapy expenditures on individual instrument items and use the estimated dollar coefficients on individual items to explain therapy expenditures. In other words, an expenditure-weighted rather than psychometrically weighted scale could be created. Because an expenditure-weighted scale is constructed to maximize expenditure prediction, it might perform better in this regard than a scale formed to satisfy psychometric criteria.

3) *Develop functional status and other scales that are less subject to ceiling effects in the outpatient therapy population.*

Some of the scales analyzed in this report were subject to strong ceiling effects (a large proportion of the sample scored as not impaired). Effort could be put into trying to distinguish degree of functional impairment in this population, possibly requiring new assessment items or at least more detailed response options to the current items.

4) *Further explore the Continuity Assessment Record and Evaluation (CARE) function items showing misfit and the potential sources of that misfit.*

CARE-C Life Skills subscale item *answering yes/no questions about basic needs* seemed to be "noisy" (i.e., there was lack of agreement between expected and observed responses on this item) and was not recommended for subsequent resource utilization analyses. Because the remaining items in this subscale also address communication- and cognition-related tasks, we recommend that this item undergo further evaluation in future research.

Clinician-reported wheelchair items also tended to be noisy but are extremely important in the assessment of mobility in beneficiaries who use wheelchairs. Some of the unexpected responses could come from the type of wheelchair used (motorized or manual), and future research could capture the differences in responding if information on the type of wheelchair was also collected.

When assessing a set of beneficiaries who have more functional limitations (e.g., nursing facility residents), consideration should be given to using full self-report subscales (similar to those in CARE-C) rather than putting in items on several different topics (e.g., mobility, self-care, and communication) into one subscale, as was implemented for CARE-F. The CARE-F self-report Patient Ability subscale items *making yourself understood* and *telling others your basic needs* showed a lack of agreement between expected and observed responses. Also, these items did not make clinical sense in relation to other items in the subscale that concern motor functions (e.g., *moving up in bed* or *taking care of personal grooming*). Therefore, future research should consider keeping these items separate if a reduced item set is ideal for those with severe functional limitations.

As noted in ***Section 3***, the distance items used for CARE-C and CARE-F clinician-reported Mobility subscales did not allow for an in-depth patient evaluation because therapists were instructed to evaluate only the furthest distance that required little or no assistance. Hence, we were unable to determine the extent of agreement between expected and observed responses. For example, if a beneficiary needed partial assistance on the furthest distance assessed, information was not obtained on their level of independence with a shorter distance. These items need to be revisited in future work with different data collection rules. A suggested change would be an initial evaluation of the furthest distance, followed by an assessment of shorter distances until the patient would require no assistance. An initial gateway question asking if the patient needs assistance at all would shorten the instrument for those who are completely ambulatory.

6.2 Conditions and Diagnoses

Patient medical conditions and diagnoses are some of the most significant predictors of therapy utilization. Further exploration of this important category of predictors is warranted. Following are some ideas in this category:

1) *Analyze the predictive power of diagnoses on current-year claims (e.g., concurrent hierarchical coexisting conditions [HCCs] or other diagnostic classifications).*

In this report, we examined the predictive ability of prior-year (prospective) diagnoses/HCCs. Prior-year diagnoses should predict therapy undertaken to treat chronic conditions. However, therapy may also be indicated for acute or incident conditions, such as fracture or new onset low back pain. In future work, the predictive power of current-year diagnoses/HCCs to explain therapy expenditures could be examined. Current-year diagnoses should explain therapy directed toward acute or new onset conditions better than prior-year diagnoses. A current-year profile of diagnoses will have a variable temporal relationship to therapy utilization—some diagnoses "pre-dicting" therapy utilization, some "post-dicting" utilization. Although this current-year profile may raise face validity concerns, statistical relationships can be developed. Sorting out the timing of diagnoses versus therapy utilization requires a more detailed, episode approach.

2) *Add non-payment HCC diagnostic categories, such as musculoskeletal HCCs. Use administrative data to calculate non-payment HCCs and try them in the models.*

In this report, we examined the predictive performance of payment HCCs (non-payment HCCs were not available). Future work could create and examine the predictive performance of non-payment HCCs.

6.3 Other Factors Affecting Therapy Costs

In addition to patient functional status and diagnoses, other factors can help to predict therapy utilization. Following are suggestions for further research on other predictive factors:

1) *Perform additional analysis of the predictive power of prior utilization.*

In this project, we examined the predictive power of several types of prior utilization. We found that timing of previous surgeries and inpatient therapy charges were related to outpatient therapy expenditures. Additional analysis to understand these relationships better could be undertaken. Some specific analyses could be to examine the following types of prior utilization:

a) surgeries of various types (e.g., fracture repair, including inpatient and outpatient surgeries and the timing relative to the initiation of outpatient therapy),
b) therapy utilization during home health episodes and/or home health resource group classifications,
c) amount of therapy utilized during skilled nursing facility and inpatient rehabilitation facility stays immediately before initiation of outpatient services, and
d) number and length of prior hospitalizations in a fixed time frame before the initiation of outpatient services.

2) *Examine usefulness of prior-year therapy expenditures in explaining current-year expenditures.*

Prior expenditures can explain current expenditures, but they may establish undesirable incentives for lack of efficiency.

3) *Examine the content of therapy episodes, as measured by Current Procedural Terminology (CPT) codes (e.g., content of 1-day episodes, short episodes versus long episodes).*

Examining the content of episodes might help explain why certain episodes for apparently similar beneficiary diagnostic conditions and functional status are much longer than others.

4) *Perform case study analysis with a few outpatient therapy practices of the factors affecting episode length and expenditures. For example, why do some episodes on apparently similar patients last only 1 day, whereas others continue for weeks or months? Also, perform a detailed analysis of the content of episodes (e.g., instruction, training, therapy, transition from/to other care settings) and how that relates to episode length and expenditures.*

Detailed case studies might uncover additional factors that can help explain therapy episode length and expenditures.

6.4 Model Specification and Estimation Approaches

Analysis of further model specification and estimation approaches may improve the prediction of therapy utilization. Following are two ideas for further work:

1) *Estimate regression models predicting therapy expenditures decomposed into 60-day periods to simulate fixed-length renewable episode-based payment.*

If episode-based payment were to be based on, for example, 60-day fixed-length episodes, analysis could focus on predicting expenditures for that period.

2) *Estimate quantile regressions for annual expenditures using the patient self-report and clinician assessment data variables. This estimate will help us determine how to case-mix-adjust the annual therapy cap.*

In this report, we estimated quantile regressions on a sample of beneficiaries with administrative data only. Additional quantile regressions for the annual therapy expenditures cap could be estimated with the CARE data items. These analyses would tell us how the clinician assessment and patient self-report variables relate to the therapy cap versus mean expenditures.

6.5 Other Suggestions

This section includes some other suggestions for further research on therapy payment models.

1) *Analyze the G-code patient functional status modifiers that the Centers for Medicare & Medicaid Services (CMS) is currently collecting for outpatient therapy.*

CMS is currently requiring therapy practices to report data on degree of impairment of patients, using G-codes on administrative claims. These data could be analyzed to determine how well degree of impairment predicts therapy expenditures or whether the data can be used to analyze outcomes. Future changes to the G-code modifiers could include participation and mobility items because under DOTPA, these scales had the strongest explanatory power to predict payment.

2) *Analyze the CARE discharge assessments to examine outcomes of care.*

This report did not analyze the CARE discharge assessments. Analysis of these data could be conducted, for example, to try to determine certain outcomes of care and how the outcomes relate to quantity of therapy visits and expenditures incurred. This would be a first step toward a therapy payment system that rewards outcomes achieved rather than quantity of therapy provided.

3) *Simulate any applicable/current Medicare Payment Assessment Commission (MedPAC) proposals, such as lowering annual outpatient therapy caps.*

Any current policy proposals of the MedPAC could be simulated and analyzed—for example, lowering the annual therapy cap.

4) *Address further questions about high-cost beneficiaries.*

What are the conditions, functional status, and circumstances that drive the highest utilizers of therapy services? How can they be adjusted for in therapy case-mix measurement?

[This page intentionally left blank.]

7. CONCLUSIONS

The goal of this project was to develop payment alternatives for Medicare outpatient therapy payment, with a particular focus on developing case-mix adjustment methods. We examined two major payment alternatives, fee-for-service (FFS) with an annual expenditures cap and episode-based payment. We also argued that a blend of FFS and episode-based payment (*mixed payment*) may be preferable to either pure payment approach because mixed payment establishes greater incentives for efficiency than FFS payment but lesser incentives for underprovision of therapy services than episode-based payment. More members of the Developing Outpatient Therapy Payment Alternatives (DOTPA) technical expert panel (TEP) favored mixed payment over any other payment option currently under consideration in this study.

As for case-mix adjustment, clinical face validity includes a focus on diagnosis and functional status. Our project TEP rated these factors the most important to include in a therapy case-mix model, along with reason for therapy. We developed several functional status scales relevant to predicting outpatient therapy expenditures and tested their predictive ability. Also, we developed 17 exploratory physical therapy (PT) case-mix groups based on diagnosis and functional status.

Future studies of outpatient therapy case-mix would benefit from giving the following considerations the highest priority. First, much larger sample sizes than were available in the DOTPA analyses are needed for analysis of occupational therapy and speech-language pathology case-mix. Although a PT case-mix analysis was conducted under DOTPA, a more refined case-mix could be developed with a larger sample size. An oversample of high-need beneficiaries should be considered for an examination of this important subgroup. Second, the current CARE instrument is too long, complex, and redundant. For payment applications, a much shorter and simpler instrument should be developed and administered through an electronic medium. Substantial missing data and nonresponse occurred in the CARE responses. A successor instrument (or possibly a set of tailored instruments) should attempt to collect only data that are feasible and relevant for the responding therapist to provide. One way to streamline would be to create discipline-specific instruments for the CARE-C population, for which discipline-specific payment is implemented. Third, function scales appropriate to the outpatient therapy population, without the marked ceiling effects observed in the CARE scales, should be developed. Fourth, both rehabilitation and medical diagnoses should be collected, and should be clearly distinguished, on the assessment instrument.

The outpatient therapy case-mix is highly diverse. Developing case-mix adjustment for the entire population may be a difficult short-term goal to achieve. A more achievable goal is to focus on case-mix adjustment for selected cases such as those in which an adjustment is easier to define or those in subpopulations for which case-mix adjustment is particularly important. An example of the former may be outpatient therapy after joint replacement. Clinical panels and empirical evidence may be able to establish relatively strong and standardized expectations for the amount of therapy after joint replacement. A bundled payment for therapy triggered by the joint replacement could then be established. An example of the latter situation may be beneficiaries with chronic disability caused by serious medical conditions or events, who can meaningfully and demonstrably benefit from therapy and who are likely to be high users of therapy. These patients could include, for example, those with post-stroke disability, Parkinson's-related disability, other severe neurological conditions, paralysis, or amputations. Case-mix adjustment could focus on distinguishing these high-need (but relatively rare) beneficiaries from the more common musculoskeletal or pain patients without serious disability. Under this perspective, it is appropriate that the CARE instrument initially focus on measuring higher degrees of impairment to distinguish high-need beneficiaries, rather than attempt to make fine distinctions among the large number of less seriously limited musculoskeletal or pain patients.

REFERENCES

Amico, P., Dever, J.A., Scheffler, S., and Potelle, J.: Developing Outpatient Therapy Payment Alternatives: Documentation for Analytic Files. Centers for Medicare & Medicaid Services Contract No. HHSM-500—2012-00169G. RTI International, 2014a.

Amico, P., Pope, G., Adamache, W., Pardasaney, P., Deutsch, A., Gordon, T., Silver, B., and Dever, J.A.: Developing Outpatient Therapy Payment Alternatives (DOTPA): Payment Alternatives Report. Centers for Medicare & Medicaid Services Contract No. HHSM-500—2012-00169G. RTI International, 2014b.

Andres, P., Haley, S.M., and Ni, P.S.: Are patient-reported functional measures reliable in monitoring post-acute outcomes? Am. J. Phys. Med. Rehabil. 82(8):614–621, 2003.

Balanced Budget Act of 1997, Pub. L. No 105-33, 111 Stat 251, 1997.

Bond, T.G., and Fox, C.M.: Applying the Rasch Model: Fundamental Measurement in the Human Sciences. Mahwah, N.J. Lawrence Erlbaum Associates, 2001.

Centers for Medicare & Medicaid Services.: Medicare Outpatient Therapy Billing. Medicare Learning Network, 2010.

Centers for Medicare & Medicaid Services.: Medicare Benefit Policy Manual: Chapter 15, 2012. Retrieved December 27, 2012, from CMS.gov: http://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/downloads/bp102c15.pdf.

Gage, B., et al.: Post-Acute Care Payment Reform Demonstration: Final Report. Centers for Medicare & Medicaid Services Contract No. HHSM-500-2005-00029I. RTI International, 2012.

Haley, S.M., Ni, P., Coster, W.J., et al.: Agreement in functional assessment: Graphical approaches to displaying respondent effects. Am. J. Phys. Med. Rehabil. 85:747–755, 2006.

Kandilov, A., Lyda-McDonald, B., and Drozd, E.M.: Developing Outpatient Therapy Payment Alternatives (DOTPA): 2007 Utilization Report. Prepared for the Centers for Medicare & Medicaid Services by RTI International, 2009.

Kline, T., Deutsch, A., Pardasaney, P., Amico, P., Gordon, T., Pope, G., Ridenhour, J., Jette, A., Silver, B., and Dever, J.A.: Developing Outpatient Therapy Payment Alternatives (DOTPA): Measurement Report. Centers for Medicare & Medicaid Services Contract No. HHSM-500—2012-00169G. RTI International, 2014.

Linacre, J.M.: Winsteps® Rasch Measurement Computer Program. Beaverton, Oreg. 2012. http://www.winsteps.com

Lyda-McDonald, B., Drozd, E.M., and Gage, B.: Developing Outpatient Therapy Payment Alternatives (DOTPA): 2009 Utilization Report. Prepared for the Centers for Medicare & Medicaid Services by RTI International, 2012.

Merriam-Webster, Inc.: Physical medicine and rehabilitation. Retrieved October 19, 2012, from Merriam-Webster's Medical Dictionary (Medline Plus). http://www.merriam-webster.com/medlineplus/physical%20medicine%20and%20rehabilitation

Silver, B., and Dever, J.A.: Developing Outpatient Therapy Payment Alternatives (DOTPA): Phase I Final Report. Centers for Medicare & Medicaid Services Contract No. 500-2005-00029I/0012. RTI International, 2013.

Silver, B., Lyda-McDonald, B., Bachofer, H., and Gage, B.: Developing Outpatient Therapy Payment Alternatives (DOTPA): 2009 Utilization Report: Part II. Prepared for the Centers for Medicare & Medicaid Services by RTI International, 2012.

Silver, B., Lyda-McDonald, B., Bachofer, H., and Gage, B.: Developing Outpatient Therapy Payment Alternatives (DOTPA): 2010 Utilization Report. Prepared for the Centers for Medicare & Medicaid Services by RTI International, 2013.

Social Security Act of 1997. U.S. Code. Title 42 Sec. 1395x, 1997.

World Health Organization.: International Classification of Functioning, Disability and Health: ICF. Geneva, Switzerland. World Health Organization, 2001.

[This page intentionally left blank.]

APPENDIX A:
ITEMS SETS WITHIN FINAL ABILITY SUBSCALES

Appendix Tables A.1 through **A.4** contain detailed information on the final Continuity Assessment Record and Evaluation (CARE)-C and CARE-F subscales, including item content and response option information (in footnotes). The tables are ordered by patient self-reported information for both CARE-C and CARE-F, followed by subscales using clinician-reported information (CARE-C and CARE-F, respectively). A detailed discussion of the recommended changes displayed in Table 3-5 is found within the *DOTPA Measurement Report* (Kline et al., 2014); these recommendations are addressed in this final item set, and the final subscale described as follows is used in the analyses described in the *DOTPA Payment Alternatives Report* (Amico et al., 2014b).

[This page intentionally left blank.]

Appendix Table A.1
CARE-C final subscale configuration: Patient self-report

Subscale	Item content
Basic Mobility[1,2]	How much help from another person do you currently need moving to and from a bed to a chair? How much difficulty do you currently have… a) Moving from sitting at the side of the bed to lying down on your back? b) Moving up in bed? c) Standing for at least 1 minute? d) Sitting down in an armless straight chair? e) Standing up from an armless straight chair? f) Getting into and out of a car/taxi? g) Walking around on one floor, taking into consideration thresholds, doors, furniture, and a variety of floor coverings? h) Going up and down a flight of stairs inside, using a handrail? i) Bending over from a standing position to pick up a piece of clothing from the floor without holding onto anything? j) Walking several blocks? k) Walking up and down steep unpaved inclines? l) Carrying something in both arms while climbing a flight of stairs?
Wheelchair Use[1]	Without help from another person, when you are using your wheelchair, how much difficulty do you currently have… a) Moving around within one room, including making turns in a wheelchair? b) Opening a door away from a wheelchair? c) Opening a door toward a wheelchair? d) Transferring between a wheelchair and other seating surfaces, such as a chair or bed? e) Propelling/driving a wheelchair several blocks?
Everyday Activities[1,2]	How much help do you currently need… a) Taking care of your personal grooming such as brushing teeth, combing hair, etc.? b) Bathing yourself? How much difficulty do you currently have… a) Picking up thin, flat objects from a table? b) Putting on and taking off a shirt or blouse? c) Putting on and taking off socks? d) Opening small containers like aspirin or vitamins?

(continued)

Appendix Table A.1 (continued)
CARE-C final subscale configuration: Patient self-report

Subscale	Item content
Everyday Activities[1,2]	How much help do you currently need… a) Removing stiff plastic packaging using hands and scissors? b) Tying shoes? c) Unscrewing the lid off a previously unopened jar without using devices? d) Washing indoor windows? e) Lifting 25 pounds from the ground to a table? f) Cutting your toenails?
Life Skills[1]	How much difficulty do you currently have… a) Understanding instructions involving several steps? b) Following/understanding a 10- to 15-minute speech or presentation? c) Making yourself understood to other people during ordinary conversations? d) Telling someone important information about yourself in case of emergency? e) Explaining how to do something involving several steps to another person? f) Reading and following complex instructions? g) Telling others your basic needs? h) Planning for and keeping appointments that are not part of your weekly routine? i) Reading simple material? j) Filling out a long form? k) Writing down a short message or note? l) Remembering where things were placed or put away? m) Keeping track of time? n) Putting together a shopping list of 10 to 15 items? o) Remembering a list of 4 or 5 errands without writing it down? p) Taking care of complicated tasks like managing a checking account or getting appliances fixed?
Participation[3]	Even with help or services, tell us how much you are limited in… a) Keeping your home clean and fixed up? b) Providing personal care to yourself? c) Getting groceries or other things for your home? How much are you currently limited in… a) Going to movies, plays, concerts, sporting events, museums, or similar activities?

NOTES:

CARE-C = Continuity Assessment Record and Evaluation form for patients living in the community.

[1] Response options: Unable, A Lot of Difficulty, A Little Difficulty, No Difficulty

[2] Response options: Unable, A Lot of Help Needed, A Little Help Needed, No Help Needed

[3] Response options: Extremely Limited, Very Much Limited, Somewhat Limited, A Little Limited, Not at All Limited

SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA) analytic files, 2014.

Appendix Table A.2
CARE-F final subscale configuration: Patient self-report

Subscale	Item content
Patient Ability[1,2]	How much difficulty do you currently have… a) Moving from sitting at the side of the bed to lying down on your back? b) Moving up in bed? c) Walking around on one floor, taking into consideration thresholds, doors, furniture, and a variety of floor coverings? How much HELP from another person do you currently need… a) Moving to and from a toilet?
Wheelchair Use[1]	Without help from another person, when you are using your wheelchair, how much difficulty do you currently have… a) Moving around within one room, including making turns in a wheelchair? b) Opening a door away from a wheelchair? c) Opening a door toward a wheelchair? d) Transferring between a wheelchair and other seating surfaces, such as a chair or bed?

NOTES:

CARE-F = Continuity Assessment Record and Evaluation form for patients who receive Part B therapy while residing in a nursing facility or receiving day rehabilitation services.

[1] Response options: Unable, A Lot of Difficulty, A Little Difficulty, No Difficulty

[2] Response options: Unable, A Lot of Help Needed, A Little Help Needed, No Help Needed

SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA), 2014.

Appendix Table A.3
CARE-C final subscale configuration: Clinician observation

Subscale	Item content (category)
Self-Care[1]	▪ Oral hygiene ▪ Wash upper body ▪ Upper-body dressing ▪ Lower-body dressing ▪ Putting on/taking off footwear
Mobility[1]	▪ Sit to lying ▪ Roll left and right ▪ Lying to sitting on side of bed ▪ Sit to stand ▪ Chair/bed-to-chair transfer ▪ Picking up object while standing (restricted to responses from the walking sample) ▪ Walk 50 ft. with two turns ▪ Walking 10 ft. on uneven surfaces ▪ Four steps ▪ Twelve steps ▪ Wheel up and down ramp ▪ Walk 500 ft., walk 150 ft., walk 50 ft., walk in room once standing – Combined into a single variable using responses from each item ▪ Wheel 500 ft., wheel 150 ft., wheel 50 ft., wheel in room once seated – Combined into a single variable using responses from each item
Instrumental Activities of Daily Living[1]	▪ Medication management–oral ▪ Make a light meal ▪ Wipe down surface and clean the cloth
Cognition	
Problem Solving[2]	Simple Problems: Following basic schedules; requesting assistance; using a call bell; identifying basic wants/needs; preparing a simple cold meal without assistance. Complex Problems: Working on a computer; managing personal, medical, and financial affairs; preparing a complex hot meal; grocery shopping; route finding and map reading without assistance.

(continued)

Appendix Table A.3 (continued)
CARE-C final subscale configuration: Clinician observation

Subscale	Item content (category)
Memory[2]	Basic Information: Personal information (e.g., family members, biographical information, physical location); basic schedules; names of familiar staff; location of therapy area without assistance. Complex Information: Complex and novel information (e.g., carry out multiple-step activities, follow a plan); anticipate future events (e.g., keeping appointments) without assistance.
Attention[2]	Simple Activities: Following simple directions; reading environmental signs or short newspaper/magazine/book passages; eating a meal; completing personal hygiene; dressing without assistance. Complex Activities: Watching a news program; reading a book; planning and preparing a meal; managing one's own medical, financial, and personal affairs without assistance.
Communication	
Spoken-Language Comprehension[2]	Basic Information: Simple 1-step directions; simple yes/no questions; simple words or short phrases; conversations about routine daily activities without assistance. Complex Information: Complex sentences, 2–3 step directions, 2–3 part messages, and a variety of complex topics without assistance.
Spoken-Language Expression[2]	Basic Information: Basic information regarding wants/needs or daily routines; using 1–2 words or short phrases without assistance. Complex Information: Thoughts/ideas using sentences; in conversations about routine daily activities or a variety of topics without assistance.
Motor Speech Production[2]	Intelligible in Short Utterances: Spontaneous production of automatic words, predictable single words, or short phrases in conversation without assistance. Intelligible in Longer Utterances: Spontaneous production of multisyllabic words in sentences without assistance.
Functional Voice[2]	Low Vocal Demand: Speaking softly; speaking in quiet environments; talking for short periods of time without assistance. High Vocal Demand: Speaking loudly; speaking in noisy environments; talking for extended periods of time without assistance.

NOTES:

CARE-C = Continuity Assessment Record and Evaluation form for patients living in the community.

[1] Response options: Dependent, Substantial/Maximal Assistance, Partial/Moderate Assistance, Supervising/Touching Assistance, Set-up/Clean-up Assistance, Independent

[2] Response options: Never/Rarely, Sometimes, Usually, Always

SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA), 2014.

Appendix Table A.4
CARE-F final subscale configuration: Clinician observation

Subscale	Item content (category)
Self-Care[1]	▪ Eating ▪ Oral hygiene ▪ Toileting hygiene ▪ Wash upper body ▪ Shower/bathe self ▪ Upper-body dressing ▪ Lower-body dressing ▪ Putting on/taking off footwear
Mobility[1]	▪ Sit to lying ▪ Roll left and right ▪ Lying to sitting on side of bed ▪ Sit to stand ▪ Chair/bed-to-chair transfer ▪ Picking up object while standing (restricted to responses from the walking sample) ▪ Walk 50 ft. with two turns ▪ Walking 10 ft. on uneven surfaces ▪ One step (curb) ▪ Four steps ▪ Twelve steps ▪ Wheel up and down ramp ▪ Walk 150 ft., walk 100 ft., walk 50 ft., walk in room once standing – Combined into a single variable using responses from each item ▪ Wheel 150 ft., wheel 100 ft., wheel 50 ft., wheel in room once seated – Combined into a single variable using responses from each item ▪ Car transfer (from instrumental activities of daily living)
Instrumental Activities of Daily Living[1]	▪ Telephone–answering ▪ Telephone–placing call ▪ Medication management–oral ▪ Medication management–inhalant/mist ▪ Medication management–injectable ▪ Make a light meal ▪ Wipe down surface and clean the cloth
Cognition	
Problem Solving[2]	Simple Problems: Following basic schedules; requesting assistance; using a call bell; identifying basic wants/needs; preparing a simple cold meal without assistance. Complex Problems: Working on a computer; managing personal, medical, and financial affairs; preparing a complex hot meal; grocery shopping; route finding and map reading without assistance.

(continued)

Appendix Table A.4 (continued)
CARE-F final subscale configuration: Clinician observation

Subscale	Item content (category)
Memory[2]	Complex Information: Complex and novel information (e.g., carry out multiple-step activities, follow a plan); anticipate future events (e.g., keeping appointments) without assistance. Basic Information: Personal information (e.g., family members, biographical information, physical location); basic schedules; names of familiar staff; location of therapy area with assistance.
Attention[2]	Simple Activities: Following simple directions; reading environmental signs or short newspaper/magazine/book passages; eating a meal; completing personal hygiene; dressing without assistance. Complex Activities: Watching a news program; reading a book; planning and preparing a meal; managing one's own medical, financial, and personal affairs without assistance.
Communication	
Spoken-Language Comprehension[2]	Basic Information: Simple 1-step directions; simple yes/no questions; simple words or short phrases; conversations about routine daily activities without assistance. Complex Information: Complex sentences, 2–3 step directions, 2–3 part messages, and a variety of complex topics without assistance.
Spoken-Language Expression[2]	Basic Information: Basic information regarding wants/needs or daily routines; using 1–2 words or short phrases without assistance. Complex Information: Thoughts/ideas using sentences; in conversations about routine daily activities or a variety of topics without assistance.
Motor Speech Production[2]	Intelligible in Short Utterances: Spontaneous production of automatic words, predictable single words, or short phrases in conversation without assistance. Intelligible in Longer Utterances: Spontaneous production of multisyllabic words in sentences without assistance.
Functional Voice[2]	Low Vocal Demand: Speaking softly; speaking in quiet environments; talking for short periods of time without assistance. High Vocal Demand: Speaking loudly; speaking in noisy environments; talking for extended periods of time without assistance.

NOTES:

CARE-F = Continuity Assessment Record and Evaluation form for patients who receive Part B therapy while residing in a nursing facility or receiving day rehabilitation services.

[1] Response options: Dependent, Substantial/Maximal Assistance, Partial/Moderate Assistance, Supervising/Touching Assistance, Set-up/Clean-up Assistance, Independent

[2] Response options: Never/Rarely, Sometimes, Usually, Always

SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA), 2014.

APPENDIX B:
RASCH TO RAW LOOK-UP TABLES

Tables B-1–B-13 in this appendix provide a brief way to approximate a Rasch measure (used in the assessment of resource use) from the data on the Continuity Assessment Record and Evaluation (CARE)-C and CARE-F item sets. These tables provide a raw score value that is created by adding the item scores for each subscale (e.g., self-report Basic Mobility). However, these tables can only be used when *all* the scores for a subscale are reported on the CARE item set; in other words, no item in the scale has missing data. In addition, the creation of several subscales was modified during the course of analysis (see ***Table 3-5***), and the correspondence between the Rasch measure and raw score is specific to these changes. Finally, the Rasch measures were rescaled to a value between 0 and 100 through a simple algebraic conversion for ease in interpreting model results.

Table B-1
CARE-C patient-reported basic mobility: Raw scale score compared with Rasch measure and standard error

Raw score	Rasch measure	Rasch SE	Raw score	Rasch measure	Rasch SE
0	0.00 (E)	13.55 (E)	20	50.15	3.08
1	10.13	8.06	21	51.51	3.10
2	16.88	6.02	22	52.87	3.11
3	21.15	5.05	23	54.25	3.14
4	24.35	4.49	24	55.66	3.16
5	26.96	4.12	25	57.09	3.20
6	29.20	3.86	26	58.56	3.24
7	31.20	3.66	27	60.07	3.29
8	33.02	3.52	28	61.63	3.36
9	34.71	3.40	29	63.27	3.44
10	36.31	3.31	30	64.99	3.54
11	37.83	3.24	31	66.83	3.66
12	39.29	3.19	32	68.80	3.82
13	40.70	3.14	33	70.97	4.01
14	42.09	3.11	34	73.39	4.27
15	43.45	3.09	35	76.17	4.62
16	44.80	3.08	36	79.51	5.12
17	46.14	3.07	37	83.81	5.96
18	47.47	3.07	38	90.29	7.85
19	48.81	3.07	39	100.00 (E)	13.37 (E)

NOTES: CARE-C = Continuity Assessment Record and Evaluation form for patients living in the community; E = value estimated because item values were missing; and SE = standard error.

SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA) analytic files, 2014.

Table B-2
CARE-C patient-reported wheelchair use: Raw scale score compared with Rasch measure and standard error

Raw score	Rasch measure	Rasch SE	Raw score	Rasch measure	Rasch SE
0	0.00 (E)	18.53 (E)	8	47.94	6.58
1	12.89	10.61	9	52.43	6.82
2	21.19	8.03	10	57.29	7.11
3	26.80	7.04	11	62.62	7.49
4	31.42	6.58	12	68.62	8.02
5	35.61	6.38	13	75.79	8.99
6	39.65	6.34	14	85.82	11.42
7	43.71	6.41	15	100.00 (E)	19.03 (E)

NOTES: CARE-C = Continuity Assessment Record and Evaluation form for patients living in the community; E = value estimated because item values were missing; and SE = standard error.
SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA) analytic files, 2014.

Table B-3
CARE-C patient-reported everyday activities: Raw scale score compared with Rasch measure and standard error

Raw score	Rasch measure	Rasch SE	Raw score	Rasch measure	Rasch SE
0	0.00 (E)	16.64 (E)	19	50.15	3.45
1	11.50	9.43	20	51.50	3.48
2	18.61	6.89	21	52.87	3.52
3	23.02	5.76	22	54.27	3.57
4	26.28	5.09	23	55.71	3.62
5	28.91	4.64	24	57.20	3.69
6	31.13	4.31	25	58.75	3.76
7	33.10	4.08	26	60.37	3.86
8	34.87	3.90	27	62.08	3.97
9	36.50	3.76	28	63.90	4.11
10	38.04	3.65	29	65.86	4.28
11	39.49	3.57	30	68.02	4.51
12	40.90	3.51	31	70.44	4.82
13	42.26	3.47	32	73.26	5.25
14	43.59	3.44	33	76.71	5.89
15	44.91	3.42	34	81.27	6.98
16	46.21	3.42	35	88.48	9.46
17	47.52	3.42	36	100.00 (E)	16.64 (E)
18	48.83	3.43	—	—	—

NOTES: — = no data are present; CARE-C = Continuity Assessment Record and Evaluation form for patients living in the community; E = value estimated because item values were missing; and SE = standard error.
SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA) analytic files, 2014.

Table B-4
CARE-C patient-reported life skills: Raw scale score compared with Rasch measure and standard error

Raw score	Rasch measure	Rasch SE	Raw score	Rasch measure	Rasch SE
0	0.00 (E)	15.28 (E)	25	50.56	2.84
1	10.45	8.60	26	51.54	2.85
2	16.85	6.29	27	52.53	2.87
3	20.85	5.28	28	53.53	2.89
4	23.83	4.69	29	54.55	2.91
5	26.26	4.28	30	55.59	2.94
6	28.32	3.99	31	56.65	2.98
7	30.13	3.76	32	57.74	3.02
8	31.77	3.59	33	58.86	3.06
9	33.26	3.44	34	60.01	3.12
10	34.65	3.33	35	61.21	3.18
11	35.95	3.23	36	62.46	3.25
12	37.18	3.15	37	63.78	3.33
13	38.36	3.08	38	65.16	3.43
14	39.49	3.03	39	66.63	3.54
15	40.58	2.98	40	68.20	3.68
16	41.64	2.94	41	69.91	3.84
17	42.68	2.91	42	71.79	4.05
18	43.69	2.89	43	73.91	4.32
19	44.69	2.87	44	76.36	4.70
20	45.68	2.85	45	79.35	5.27
21	46.66	2.84	46	83.30	6.26
22	47.64	2.83	47	89.63	8.56
23	48.61	2.83	48	100.00 (E)	15.25 (E)
24	49.58	2.84	—	—	—

NOTES: — = no data are present; CARE-C = Continuity Assessment Record and Evaluation form for patients living in the community; E = value estimated because item values were missing; and SE = standard error.
SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA) analytic files, 2014.

Table B-5
CARE-C patient-reported participation: Raw scale score compared with Rasch measure and standard error

Raw score	Rasch measure	Rasch SE	Raw score	Rasch measure	Rasch SE
0	0.00 (E)	18.19 (E)	9	53.67	6.29
1	13.17	10.75	10	57.87	6.45
2	22.24	8.35	11	62.30	6.63
3	28.50	7.30	12	67.00	6.87
4	33.53	6.68	13	72.18	7.31
5	37.88	6.32	14	78.32	8.20
6	41.89	6.14	15	87.05	10.58
7	45.76	6.11	16	100.00 (E)	18.12 (E)
8	49.66	6.17	—	—	—

NOTES: — = no data are present; CARE-C = Continuity Assessment Record and Evaluation form for patients living in the community; E = value estimated because item values were missing; and SE = standard error.

SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA) analytic files, 2014.

Table B-6
CARE-C clinician-reported self-care: Raw scale score compared with Rasch measure and standard error

Raw score	Rasch measure	Rasch SE	Raw score	Rasch measure	Rasch SE
5	0.00 (E)	19.14 (E)	18	55.7	4.80
6	13.07	10.87	19	57.89	4.74
7	21.41	8.18	20	60.03	4.70
8	26.93	7.07	21	62.15	4.69
9	31.27	6.42	22	64.27	4.71
10	34.95	5.97	23	66.44	4.79
11	38.19	5.65	24	68.70	4.92
12	41.12	5.40	25	71.13	5.14
13	43.83	5.22	26	73.83	5.49
14	46.38	5.08	27	77.02	6.08
15	48.82	4.99	28	81.19	7.21
16	51.17	4.91	29	88.01	10.09
17	53.46	4.85	30	100.00 (E)	18.73 (E)

NOTES: CARE-C = Continuity Assessment Record and Evaluation form for patients living in the community;
E = value estimated because item values were missing; and SE = standard error.

SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA) analytic files, 2014.

Table B-7
CARE-C Clinician-reported mobility: Raw scale score compared with Rasch measure and standard error

Raw score	Rasch measure	Rasch SE	Raw score	Rasch measure	Rasch SE
13	0.00 (E)	20.51 (E)	46	48.52	2.73
14	12.14	10.54	47	49.15	2.73
15	18.4	7.16	48	49.78	2.73
16	21.86	5.75	49	50.41	2.74
17	24.27	4.96	50	51.05	2.74
18	26.14	4.45	51	51.69	2.75
19	27.68	4.09	52	52.33	2.77
20	29.00	3.83	53	52.99	2.78
21	30.18	3.63	54	53.65	2.81
22	31.24	3.47	55	54.32	2.83
23	32.22	3.34	56	55.01	2.86
24	33.14	3.24	57	55.71	2.89
25	34.01	3.16	58	56.43	2.93
26	34.84	3.09	59	57.17	2.98
27	35.64	3.04	60	57.93	3.03
28	36.41	2.99	61	58.72	3.09
29	37.16	2.96	62	59.55	3.15
30	37.89	2.93	63	60.40	3.22
31	38.61	2.90	64	61.30	3.30
32	39.31	2.88	65	62.25	3.38
33	40.01	2.86	66	63.24	3.48
34	40.70	2.84	67	64.30	3.58
35	41.38	2.83	68	65.42	3.71
36	42.05	2.81	69	66.63	3.85
37	42.72	2.80	70	67.94	4.02
38	43.38	2.79	71	69.37	4.22
39	44.04	2.78	72	70.97	4.48
40	44.69	2.77	73	72.80	4.82
41	45.33	2.76	74	74.96	5.31
42	45.98	2.75	75	77.67	6.06
43	46.62	2.74	76	81.43	7.39
44	47.25	2.74	77	87.91	10.62
45	47.89	2.73	78	100.00 (E)	20.37 (E)

NOTES: CARE-C = Continuity Assessment Record and Evaluation form for patients living in the community; E = value estimated because item values were missing; and SE = standard error.
SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA) analytic files, 2014.

Table B-8
CARE-C clinician-reported instrumental activities of daily living: Raw scale score compared with Rasch measure and standard error

Raw score	Rasch measure	Rasch SE	Raw score	Rasch measure	Rasch SE
3	0.00 (E)	26.65 (E)	11	47.15	6.92
4	15.83	14.10	12	50.30	7.12
5	24.71	10.08	13	53.71	7.50
6	30.15	8.50	14	57.61	8.14
7	34.30	7.67	15	62.41	9.25
8	37.83	7.21	16	69.05	11.29
9	41.04	6.96	17	80.43	16.00
10	44.10	6.87	18	100.00 (E)	28.86 (E)

NOTES: CARE-C = Continuity Assessment Record and Evaluation form for patients living in the community; E = value estimated because item values were missing; and SE = standard error.
SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA) analytic files, 2014.

Table B-9
CARE-F patient self-reported ability: Raw scale score compared with Rasch measure and standard error

Raw score	Rasch measure	Rasch SE	Raw score	Rasch measure	Rasch SE
0	0.00 (E)	14.36 (E)	7	58.37	7.11
1	11.04	8.94	8	65.05	7.01
2	19.99	7.80	9	71.59	7.02
3	28.20	7.93	10	78.60	7.58
4	36.66	7.88	11	87.93	9.41
5	44.49	7.42	12	100.00 (E)	14.78 (E)
6	51.56	7.17	—	—	—

NOTES: — = no data are present; CARE-F = Continuity Assessment Record and Evaluation form for patients who receive Part B therapy while residing in a nursing facility; E = value estimated because item values were missing; and SE = standard error.
SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA) analytic files, 2014.

Table B-10
CARE-F patient self-reported wheelchair use: Raw scale score compared with Rasch measure and standard error

Raw score	Rasch measure	Rasch SE	Raw score	Rasch measure	Rasch SE
0	0.00 (E)	15.71 (E)	7	59.83	7.10
1	12.46	9.86	8	66.11	7.16
2	22.17	8.19	9	72.65	7.37
3	30.10	7.93	10	79.71	7.78
4	38.12	8.12	11	88.46	9.36
5	46.18	7.88	12	100.00 (E)	15.35 (E)
6	53.38	7.36	—	—	—

NOTES: — = no data are present; CARE-F = Continuity Assessment Record and Evaluation form for patients who receive Part B therapy while residing in a nursing facility; E = value estimated because item values were missing; and SE = standard error.
SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA) analytic files, 2014.

Table B-11
CARE-F clinician-reported self-care: Raw scale score compared with Rasch measure and standard error

Raw score	Rasch measure	Rasch SE	Raw score	Rasch measure	Rasch SE
8	0.00 (E)	14.44 (E)	29	55.18	3.17
9	10.61	8.53	30	56.50	3.16
10	17.70	6.45	31	57.81	3.16
11	22.33	5.49	32	59.12	3.15
12	25.86	4.90	33	60.43	3.16
13	28.76	4.51	34	61.75	3.17
14	31.26	4.22	35	63.08	3.19
15	33.48	4.00	36	64.43	3.21
16	35.50	3.83	37	65.81	3.26
17	37.37	3.70	38	67.23	3.31
18	39.13	3.6	39	68.70	3.39
19	40.80	3.52	40	70.26	3.49
20	42.40	3.45	41	71.92	3.62
21	43.94	3.40	42	73.73	3.80
22	45.44	3.35	43	75.74	4.03
23	46.91	3.31	44	78.06	4.37
24	48.34	3.28	45	80.86	4.89
25	49.75	3.25	46	84.55	5.79
26	51.13	3.23	47	90.43	7.89
27	52.50	3.21	48	100.00 (E)	14.03 (E)
28	53.84	3.19	—	—	—

NOTES: — = no data are present; CARE-F = Continuity Assessment Record and Evaluation form for patients who receive Part B therapy while residing in a nursing facility; E = value estimated because item values were missing; and SE = standard error.
SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA) analytic files, 2014.

Table B-12
CARE-F clinician-reported mobility: Raw scale score compared with Rasch measure and standard error

Raw score	Rasch measure	Rasch SE	Raw score	Rasch measure	Rasch SE
15	0.00 (E)	13.91 (E)	45	50.06	2.54
16	9.43	7.79	46	50.92	2.53
17	15.14	5.67	47	51.76	2.53
18	18.68	4.76	48	52.61	2.52
19	21.34	4.23	49	53.44	2.51
20	23.51	3.88	50	54.27	2.50
21	25.37	3.63	51	55.09	2.49
22	27.02	3.43	52	55.91	2.47
23	28.51	3.28	53	56.71	2.46
24	29.88	3.16	54	57.51	2.45
25	31.16	3.06	55	58.30	2.44
26	32.37	2.98	56	59.08	2.43
27	33.51	2.90	57	59.86	2.42
28	34.61	2.84	58	60.63	2.41
29	35.66	2.79	59	61.40	2.40
30	36.67	2.75	60	62.16	2.40
31	37.66	2.71	61	62.92	2.39
32	38.62	2.68	62	63.67	2.39
33	39.55	2.65	63	64.43	2.38
34	40.47	2.62	64	65.18	2.38
35	41.38	2.61	65	65.93	2.38
36	42.27	2.59	66	66.68	2.38
37	43.16	2.58	67	67.43	2.37
38	44.03	2.57	68	68.17	2.37
39	44.91	2.56	69	68.92	2.37
40	45.77	2.55	70	69.66	2.37
41	46.63	2.55	71	70.41	2.37
42	47.50	2.55	72	71.15	2.37
43	48.35	2.54	73	71.90	2.37
44	49.21	2.54	74	72.65	2.38
75	73.40	2.39	83	80.07	2.77
76	74.15	2.40	84	81.14	2.92
77	74.92	2.42	85	82.35	3.13
78	75.70	2.44	86	83.76	3.43
79	76.50	2.47	87	85.53	3.91
80	77.32	2.52	88	87.97	4.76
81	78.18	2.58	89	92.18	6.84
82	79.09	2.66	90	100.00 (E)	13.08 (E)

NOTES: CARE-F = Continuity Assessment Record and Evaluation form for patients who receive Part B therapy while residing in a nursing facility; E = value estimated because item values were missing; and SE = standard error.
SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA) analytic files, 2014.

Table B-13
CARE-F clinician-reported instrumental activities of daily living: Raw scale score compared with Rasch measure and standard error

Raw score	Rasch measure	Rasch SE	Raw score	Rasch measure	Rasch SE
7	0.00 (E)	19.84 (E)	25	54.59	3.83
8	12.63	10.72	26	55.89	3.76
9	19.85	7.66	27	57.15	3.72
10	24.25	6.43	28	58.39	3.69
11	27.58	5.77	29	59.62	3.68
12	30.37	5.38	30	60.84	3.70
13	32.86	5.14	31	62.09	3.73
14	35.17	4.98	32	63.36	3.79
15	37.36	4.88	33	64.68	3.88
16	39.46	4.79	34	66.08	4.00
17	41.50	4.70	35	67.59	4.18
18	43.45	4.61	36	69.25	4.42
19	45.32	4.50	37	71.15	4.76
20	47.10	4.37	38	73.40	5.26
21	48.77	4.24	39	76.25	6.04
22	50.34	4.12	40	80.27	7.43
23	51.83	4.01	41	87.28	10.68
24	53.24	3.91	42	100.00 (E)	19.96 (E)

NOTES: CARE-F = Continuity Assessment Record and Evaluation form for patients who receive Part B therapy while residing in a nursing facility; E = value estimated because item values were missing; and SE = standard error.
SOURCE: Developing Outpatient Therapy Payment Alternatives (DOTPA) analytic files, 2014.