Research Report

Development of a Model for the Validation of Work Relative Value Units for the Medicare Physician Fee Schedule

Barbara O. Wynn, Lane F. Burgette, Andrew W. Mulcahy, Edward N. Okeke,

Ian Brantley, Neema Iyer, Teague Ruder, Ateev Mehrotra

Sponsored by the Centers for Medicare & Medicaid Services

RAND CORPORATION

For more information on this publication, visit www.rand.org/t/rr662

Published by the RAND Corporation, Santa Monica, Calif.

Preface

The Centers for Medicare & Medicaid Services (CMS) uses the resource-based relative value system (RBRVS) to pay physicians and other practitioners for their professional services. The relative values for physician work measure the relative levels of professional time, effort, skill, and stress associated with providing services. CMS asked RAND to develop a model to validate the physician work values using external data sources.

This project was designed to investigate the feasibility of developing a model and the methodological issues and limitations involved in doing so. The results presented in this report should be considered exploratory analyses that examine the overall feasibility of the model and the sensitivity of the model results to alternative methodological approaches and assumptions. The findings should be of interest to health policymakers, representatives of physician and other practitioner professional associations, and health services researchers.

This study was funded by the Centers for Medicare & Medicaid Services under contract HHSM-500-2012-00163G with Chava Sheffield as project officer. This research was conducted in RAND Health, a division of the RAND Corporation. A profile of RAND Health, abstracts of its publications, and ordering information can be found at http://www.rand.org/health.

Table of Contents

Figures

Tables

Summary

The Centers for Medicare & Medicaid Services (CMS) uses the resource-based relative value scale (RBRVS) to pay physicians and nonphysician practitioners for their professional services.[1] Under RBRVS, payment for a specific service is broken into three elements: physician work, practice expense, and malpractice expense. Concerns have been raised about the current process used by CMS to value physician work. Section 3134 of the Affordable Care Act required that CMS establish a process to validate the physician work associated with medical services. CMS asked RAND to develop a validation model for physician work values. This project was designed to describe the methodological issues and limitations involved in developing such a model.

The RBRVS system provides a total work relative value unit (RVU) for each procedure. The total work RVU for a procedure is composed of four components: (1) pre-service work (for example, positioning prior to surgery), intra-service work (the performance of the procedure or "skin-to-skin" time), (3) immediate post-service (for example, management of a patient in the post-operative period), and (4) post-operative evaluation and management (E&M) visits (only applicable for surgical procedures paid on a global period). One can calculate total work RVUs by summing each of the four components together, which has been termed the building block method (BBM), as illustrated in the following formula.

$$\begin{aligned}\text{Total Work RVUs} &= \text{Pre-service work} + \text{Intra-service work} + \text{Immediate post-service work} \\ &+ \text{Post-operative E\&M visit work}\end{aligned}$$

Each of these four work components can be broken down further as a function of time and intensity. For example, intra-service work can be divided into intra-service time and intra-service work per unit time (IWPUT) (intensity). This is illustrated in the following two formulas.

$$\text{Intra-service work} = \text{Intra-service time} \times \text{IWPUT}$$

$$\text{IWPUT} = \frac{\text{Intra-service work}}{\text{Intra-service time}}$$

[1] For simplicity, we use the terms "physician fee schedule" and "physician" throughout this report. However, the fee schedule also applies to Part B covered services furnished by certain other practitioners under their scope of practice—for example, nurse practitioners, clinical social workers, clinical psychologists, physical therapists, and others.

RAND's goal in this project is to test the feasibility of using data from external data sources and regression analysis to create prediction models to validate work RVUs. We believe the RAND model estimates could be used for two key applications. First, CMS could flag codes as potentially misvalued if the CMS and RAND model estimates are notably different. Second, CMS could also use the RAND estimates as an independent estimate of the work RVUs to consider when assessing a RUC recommendation. In some cases, further review will identify a clinical rationale for why a code is valued differently than the RAND model predictions and the CMS estimate or RUC recommendation is appropriate. In other cases, the RAND validation model results will highlight that the code was not valued accurately.

The data sets that are available to us for this project have data on intra-service time for surgical and selected medical procedures, such as interventional cardiology procedures, that are provided in hospital inpatient and outpatient settings and in ambulatory surgical centers. Our analyses focus on approximately 3,000 procedures that are often performed in an operating room setting.

Figure S.1. Overview of Modeling Approach

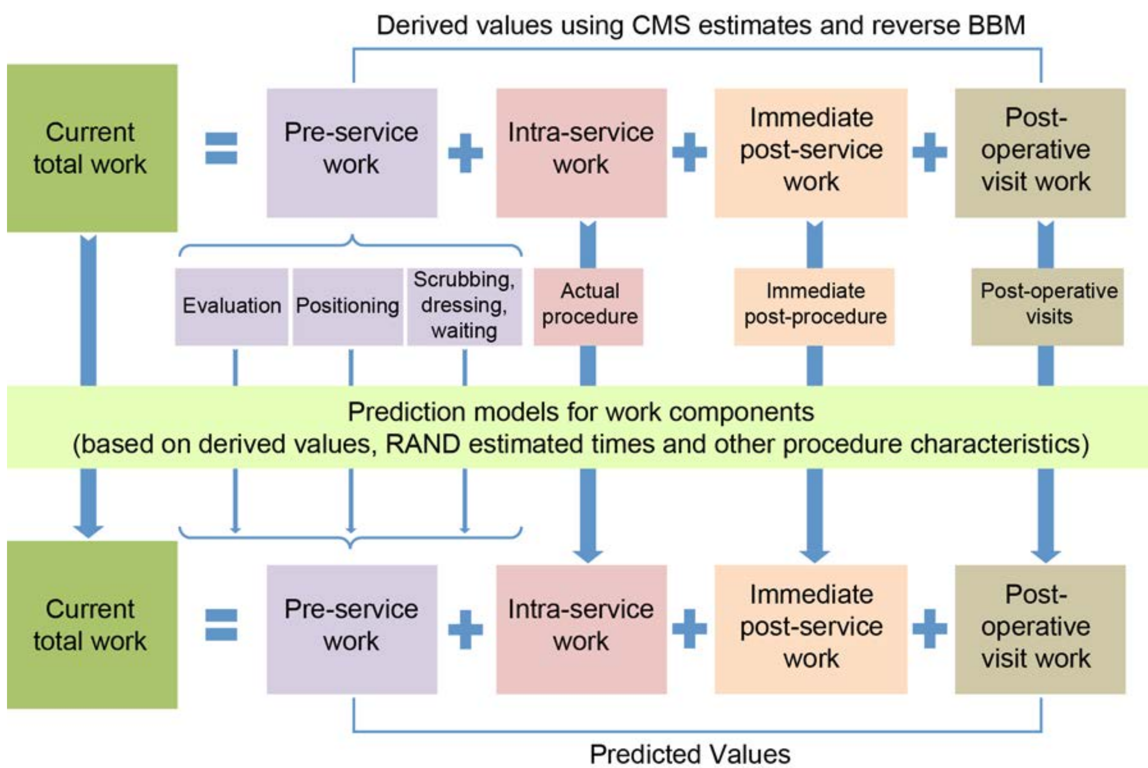


Figure S.1 provides an overview of the three overall steps in our modeling process. First, using current CMS estimates and a "reverse" BBM, we calculate for each of the four components the work, time, and intensity associated with performing each procedure. When applicable, we make corrections to these values when they do not have face validity. For example, we make a correction for procedures with negative or implausibly low intra-service work.

Second, using data from CMS and other sources, we measure 16 characteristics of the procedure. These include mortality rate after the procedure, the least-resource-intensive setting in which Medicare covers the service (i.e., inpatient, outpatient, ambulatory surgical center, office), the setting in which the procedure is typically or most often performed, and average years of training among practitioners who perform the procedure. One key characteristic that we capture using external databases is the intra-service time.

Third, we use regression analysis to build prediction models for the four work components. We use the results from the regression analyses to estimate values for each procedure and then use the BBM to combine the values for each work component into an estimate for total work RVUs. We also use a single prediction model to predict total work directly. The modeling process is complex, and we are cognizant that there are many options we could pursue at different steps. For example, should the models reflect the place of service where the procedure is typically performed, or should they reflect all the places of service where the procedure is performed? How should changes in the time required to perform a procedure affect intra-service work? To understand the impact of these methodological issues, we have created prediction models that reflect different choices:

Model 1 estimates total work for the "typical" setting in which the procedure is performed using the BBM.

> Model 1a assumes that changes in intra-service times do not affect intra-service work—i.e., changes in intra-service time are offset by changes in intensity.
> Model 1b is a blend of Model 1a and Model 1c.
> Model 1c assumes that changes in intra-service times affect intra-service work—i.e., changes in intra-service time do not affect intensity.

Model 2 estimates total work for all places of service in which the procedure is performed in our data and assumes that changes in intra-service times do not affect intra-service work.

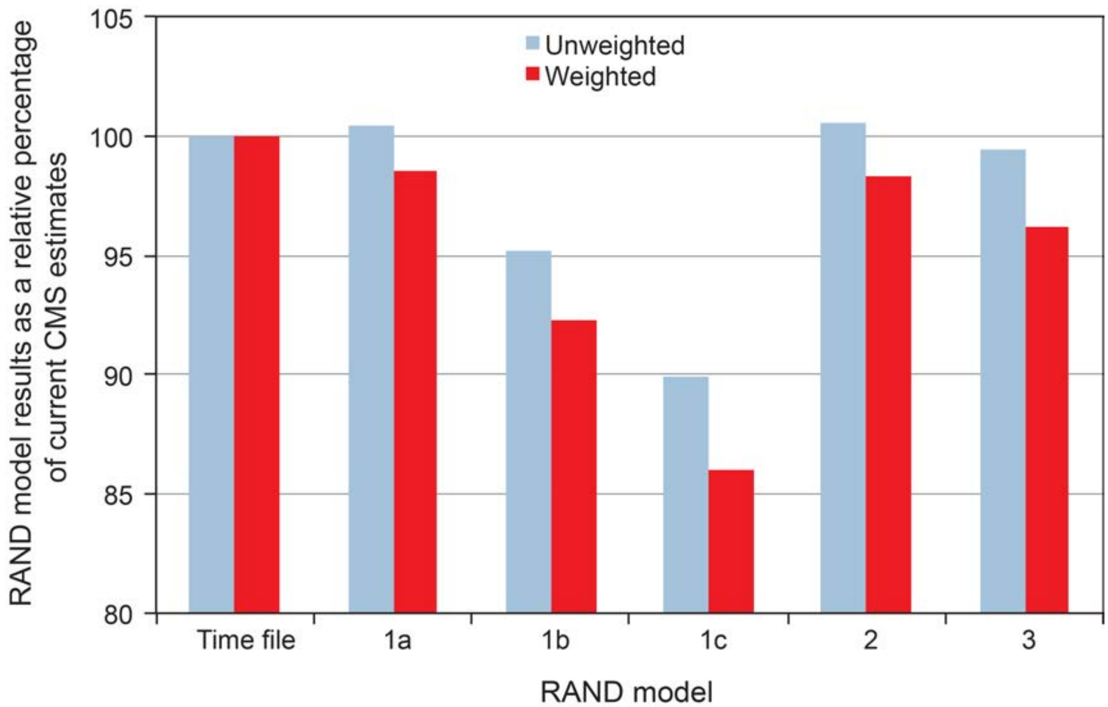
Model 3 uses a single prediction model to predict total work directly.

There were five key findings from our analyses of the RAND model estimates:

1. *RAND estimates of intra-service time using data in external datasets are typically shorter than the current CMS estimates.* For 83 percent of the procedures, the RAND time is shorter than the CMS estimates. The difference in time varies by where the procedure is performed. For example, on average, inpatient procedures are 6 percent shorter than CMS estimates, while procedures done outside the hospital with anesthesia are 20 percent shorter than CMS estimates. These results are consistent with previous research that has found that CMS time estimates are on average longer than observed times found in empirical datasets.

2. *Total work in RAND models is similar to CMS estimates, but there are important differences for some procedures.* When surgical services are unweighted by surgical volume, the average total work in RAND Models 1a, 2, and 3 and CMS total work are nearly identical, while the average RVUs for Models 1b and Model 1c are 4.8 percent and 10.0 percent lower, respectively, than the CMS average (Figure S.2). While on average the valuations of surgical procedures are similar in Model 1a, there are notable differences across the types of procedures. For example, for shorter procedures (0–30 minutes), the work estimates are 14.6 percent higher than CMS estimates, while for longer procedures (<120 minutes) the work estimates are 2.7 percent lower.

Figure S.2. Mean Total Work RVUs Predicted by Models Relative to CMS Values



3. *The difference in total RVUs across RBRVS is greater than the average difference across procedures.* The average difference between current CMS and predicted RAND values can be summarized using unweighted estimates (average difference across all procedures) or weighted estimates (the differences for high-volume procedures have a greater impact). The difference between unweighted and weighted results is important because the weighted estimates capture what would be paid by Medicare. For several models there are notable differences between the weighted and unweighted results. For example, the average total work RVUs under Model 1c as a percentage of CMS values are 90 percent and 86 percent (unweighted and weighted, respectively). There is a greater reduction in the weighted results because high-Medicare-volume procedures have higher reductions on average in the intra-service work component than low-volume procedures.

4. *Corrections reduce post-operative E&M visit work by 10 percent.* Post-operative visits on average make up 41 percent of total work among the procedures we focused on in this analysis. For a subset of procedures, we identified anomalies in the data. For example, we identified procedures for which there were inpatient E&M visits included in the global period, but the procedure is typically performed outside the hospital. These corrections reduced the unweighted average number of post-operative work RVUs by 10 percent.

5. *The difference between the CMS estimates and the RAND estimates for IWPUT and intra-service work varies across the models.* RAND estimates of intra-service time, which are based on data in external datasets, are typically shorter than the current CMS estimates. The implications of this decrease in time on IWPUT and therefore intra-service work vary under the RAND models. Under Models 1a and 2, intra-service work stays constant, intra-service time goes down, and therefore IWPUT increases. Under Model 1c, IWPUT stays the same, intra-service time is lower, and therefore intra-service work is reduced. This reduction in intra-service work is the primary reason total work under Model 1c is also reduced.

The results presented in this report should be considered exploratory analyses that examine the overall feasibility of the model and the sensitivity of the model results to alternative methodological approaches and assumptions. The final chapter of the report outlines both several important limitations of the RAND models and future analyses that could be used to address those limitations. In this chapter, we also discuss other applications of the model results and what would be necessary to expand the RAND model to nonsurgical procedures.

Acknowledgments

We would like to thank CMS staffers who provided support at various stages of this project: Kathy Bryant, Ryan Howe, Chava Sheffield, Christina Smith Ritter, Donald Thompson, and Sara Vitolo.

We also benefited from the experiences and insights of others who have been involved in valuing services under RBRVS. Our acknowledgement of their contributions does not imply that these individuals or organizations endorse the contents of this report. In particular, we would like to thank the members of our technical expert panel for their contributions at critical stages of the project: Peter Braun, Braun Consulting, PC; Allen Dobson, Dobson and DaVanzo Consultants; Iola D'Sousa, U.S. General Accountability Office; Terrence Kay, consultant; Miriam Laugesen, Columbia School and Public Health; Charles Mabry, Surgeons of South Arkansas; and Richard Whitten, Noridian Administrative Services, LLC. Their vast knowledge of how the RBRVS valuation process has evolved over time was extremely helpful in framing the issues and shaping our analyses and interpretation of results.

We also appreciate the assistance provided by the American Academy of Dermatology, the American Academy of Orthopaedic Surgery, the American Gastoenterological Association, and the American Academy of Ophthalmology in soliciting physicians for our clinical panels.

We benefited greatly from the thoughtful comments provided by Justin Dimick, Jose Escarce, Paul Ginsberg, and Susan Paddock on earlier versions of this report. Finally, we would like to thank our RAND colleagues Diane Egelhoff and Kendra Wilsher for their assistance in preparing this report and for providing general administrative support for the project and Nora Spiering and Sandy Petitjean for their assistance with this manuscript.

Abbreviations

ACA	Affordable Care Act
AMA	American Medical Association
ASC	ambulatory surgical center
BBM	building block method
CABG	coronary artery bypass graft
CMS	Centers for Medicare & Medicaid Services
CPT	Common Procedure Terminology
ED	emergency department
E&M	evaluation and management
HCPCS	Healthcare Common Procedure Coding System
HOPD	hospital outpatient department
ICD	International Classification of Diseases
IWPUT	intra-service work per unit of time
MPC	multispecialty points of comparison
NSQIP	National Surgical Quality Improvement Program
OIG	Office of Inspector General
OPPS	outpatient prospective payment system
OR	operating room
POS	place of service
RUC	American Medical Association/Specialty Society Relative Value Scale Update Committee
RVRBS	resource-based relative value scale
RVU	relative value unit

Glossary

Term	Definition
building block method (BBM)	Approach used to construct total work RVUs based on the values for the component pieces of the service. The reverse BBM approach deconstructs total work RVUs into their component parts.
CMS estimate	We use this term generically to refer to values calculated from the CMS rulemaking documents for the physician fee schedule, including Addendum B and the physician time file.
code groupings	We group codes into clinically coherent groups described in Chapter 3 and Appendix B and use the groupings in our prediction models.
code/procedure/services	These terms are generally interchangeable in this report. A service is described by a procedure code (CPT or HCPCS).
core procedure	One of the 3,179 procedures that are the focus of this project and included in our prediction models
global period	Surgical procedures generally have a 0-, 10-, or 90-day global period during which follow-up post-operative visits are bundled into the payment for the procedure. A 90-day global period also includes related visits on the day before the procedure.
Intensity	Often used synonymously with effort, it refers to the cognitive effort and judgment, technical skill and physical effort, and stress due to potential patient risk associated with delivering a service.
intra-service	The work component that accounts for the actual “face-to-face” or “skin-to-skin” time with the patient
multiple procedures	Refers to surgical encounters involving more than one surgical procedure. Medicare payments for major procedures that are furnished during the same operative session are discounted.

Term	Definition
post-service	Includes non-"skin-to-skin" work in the operating room, patient stabilization in the recovery area, and communicating with the patient/family and other professionals, as reported as a single estimate in the CMS estimates
pre-service	The work component that involves patient evaluation before a procedure, patient positioning, and scrubbing
RAND estimate	We use the term generically to refer to estimated values developed through our analyses and prediction model.
resource-based relative value scale (RBRVS)	A scaling system used in the Medicare fee schedule for physician services that measures the work done by the physicians and the costs of malpractice insurance and practice expenses relative to other services.
Time	Procedure time from CMS and RUC time estimates. There are separate estimates for intra-service, pre-service, and immediate post-service time.
typical place of service	The most common setting in which a service is provided
Volume	The number of procedures performed on Medicare beneficiaries in 2012. Volume is calculated from the utilization file produced as part of the 2014 Medicare physician fee schedule and is adjusted for multiple procedure discounts, bilateral procedures, and the use of two or more surgeons.
Work	Defined as the physician’s time and effort spent in furnishing the service. Using the BBM, it is defined as the product of time and intensity for each work component.
work components	We use this term to describe generically the four components of physician work that are used in the BBM: pre-service, intra-service, immediate post-service, and post-operative visits.

1. Introduction

Chapter Overview

In this chapter, we summarize the purpose of our project, provide an overview of how physician work is valued under the Medicare physician fee schedule, describe our project objectives and how the results might be used, and explain the organization of the remainder of the report.

Purpose

The Centers for Medicare & Medicaid Services (CMS) uses the resource-based relative value scale (RBRVS) to pay physicians and nonphysician practitioners for their professional services. Under RBRVS, payment for a specific service is broken into three elements: physician work, practice expense, and malpractice expense. Each component is valued separately in relative value units (RVUs). Total RVUs are adjusted for geographic price differences and multiplied by a dollar conversion factor to determine the Medicare fee schedule amount.

The work RVUs measure the relative levels of physician time, effort, skill, and stress associated with providing the service. Under the current process for updating the RVUs for physician work, CMS considers recommendations from the American Medical Association (AMA)/Specialty Society Relative Value Scale Update Committee (RUC) and the Medicare Payment Advisory Commission (MedPAC), public comments received through the rulemaking process, and, when warranted, recommendations from multispecialty refinement panels before establishing values for new, revised, or potentially misvalued codes.

Section 3134 of the Affordable Care Act (ACA) amended the Medicare law to require that CMS establish a process to validate the values assigned to services.[2] CMS asked RAND to develop a model to validate the physician work values of the Medicare physician fee schedule using data from existing databases that are independent of the current valuation process. The project was designed to investigate the feasibility of developing a model and the methodological issues and limitations involved in doing so. The results presented in this report should be considered exploratory analyses that examine the overall feasibility of the model and the sensitivity of the model results to alternative methodological approaches and assumptions.

[2] We define *service* to include any visit or procedure. Our use of *service* to capture the aforementioned is consistent with other published work (Braun et al., 1992; Hsiao et al., 1992; Mabry et al., 2005), though terms such as *code* or *procedure* are used in some published papers. We use the term *procedure* when discussing the surgical services that are the focus of this study.

Current System for Valuing Physician Work

In 2014, an estimated $87 billion in allowed charges was paid under the physician fee schedule for services furnished to fee-for-service Medicare beneficiaries by physicians and nonphysician practitioners (CMS, 2014a).[3] Since 1992, the physician fee schedule has been based on national RVUs that reflect the relative resources used to furnish a service. The focus of this study is the work RVUs, which account for approximately 48 percent of payments under the fee schedule. Work is defined as the physician's time spent in furnishing the service (including pre- and post-service, as well as intra-service time) and the intensity of the service. Intensity reflects the cognitive effort and judgment, technical skill and physical effort, and stress due to potential patient risk. Work values are established for most services described by the Current Procedural Terminology (CPT) coding system established and maintained by the American Medical Association (AMA, 2012).[4]

Figure 1.1 summarizes the components of total work for surgical procedures. Unlike other services, surgical procedures may include post-operative visits that occur within a global period (0, 10, or 90 days following the surgical procedure) and, in the case of surgeries with 90-day global periods, a preoperative visit occurring the day before the procedure.[5]

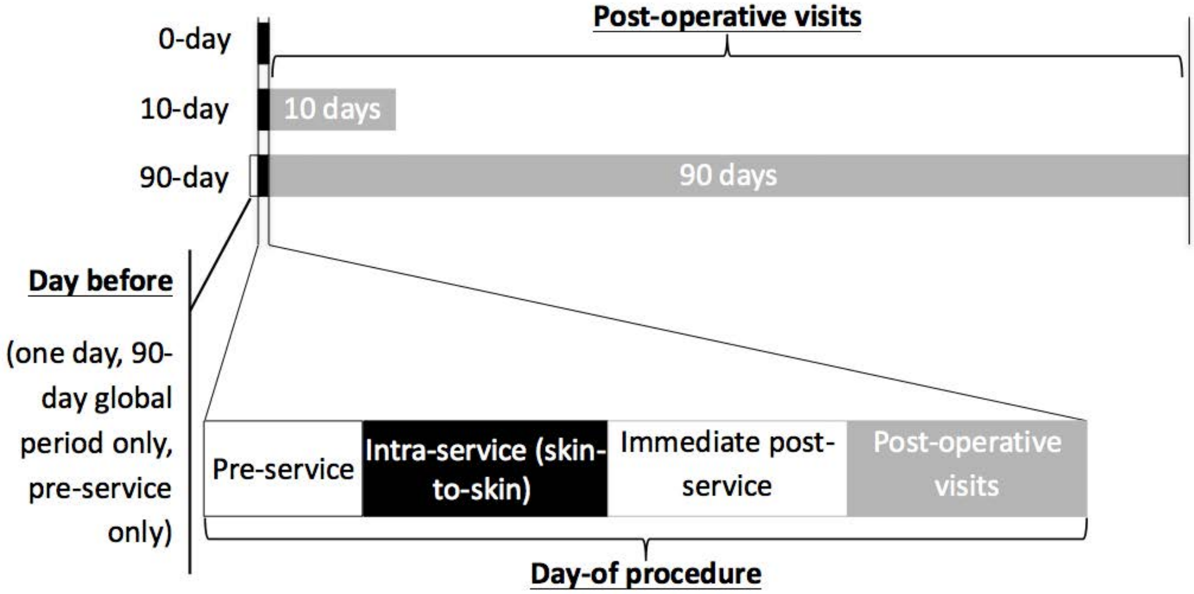
---

[3] For simplicity, we use the terms *physician fee schedule* and *physician* throughout this report. However, the fee schedule also applies to certain other practitioners for covered services provided under their scope of practice to Medicare beneficiaries—for example, nurse practitioners, clinical social workers, clinical psychologists, physical therapists, etc.

[4] Medicare uses the Healthcare Common Procedure Coding System (HCPCS). Level I is current CPT codes, and level II is alphanumeric codes assigned to services (mostly nonprofessional), medications, supplies, and equipment. CPT codes, descriptions and other data only are copyright 2013 American Medical Association. All rights reserved. Medicare uses alphanumeric codes rather than the CPT codes for a few professional services, such as new services. Services that Medicare does not cover or that are contractor-priced (such as new technology) may not have work values.

[5] In the final rule for the 2015 physician fee schedule, CMS indicated that it will begin phasing out the inclusion of post-operative E&M visits in the payment for surgical procedures that were assigned a global period in 2017 (CMS, 2014b).

Figure 1.1. Components of Total Work for Surgical Procedures



The initial work RVUs were established through a major research study led by William Hsaio at Harvard University. Key steps involved in setting the RVUs included the following (Kahan et al., 1994; Hsiao et al., 1988a; Hsiao et al., 1988b):[6]

- *Obtain specialty-specific measures of work.* Specialty-specific estimates for a set of services were obtained through telephone surveys of a nationally representative sample of physicians who were asked to use "magnitude estimation" to rate the intra-service work for all services in their survey relative to a "standard" service. Each service was described by a vignette. For example, the standard service for general surgeons was an uncomplicated indirect inguinal hernia repair on a 45-year-old male. It was assigned a value of 100, and the respondents were asked to rate other services so that a procedure requiring half the work effort received a rating of 50.
- *Link specialty-specific work measures to a common scale*. A multispecialty panel identified pairs of services from different specialties that had similar intra-service work. The linked services were combined in a regression analysis to transform the specialty-specific measurements to a common scale.

[6] Two other steps in the original methodology were to map the vignettes to CPT codes and to estimate the values for non-surveyed procedures. These steps are not directly relevant to the current update process. The original vignettes were not defined by CPT code at the outset and needed to be mapped to the CPT codes used in the fee schedules. The current update process begins with a review of the current CPT code, and definitions that require clarification or revision are referred to the CPT Editorial Board for consideration before the RUC refers the service to the specialty advisory group. The original surveys covered only a subset of services. Initially, the RVUs for the remaining services were extrapolated by using the ratio of charges within "families" of CPT procedure codes; subsequently, physician panels directly estimated the work RVUs for these procedures.

- *Add estimates of pre- and post-service work.* Estimates of pre- and post-service times were obtained for selected services. Regression models used these data to estimate pre- and post-service time for all services in the survey. These estimates were multiplied by fixed work intensity values (work per unit of time) to produce estimates of the pre- and post-service work, which were then added to the intra-service work value to create a total work value for the service.

From the outset, a process to review and revise the work RVUs was needed because continuing refinement and updating are required as CPT codes are modified, new technology is introduced, and practice patterns change. The Medicare law (Section 1848(c)(2)(B) of the Social Security Act) requires that CMS review the relative values at least every five years. Since 1991, the AMA has supported the work RVU update process through the multispecialty RUC. The RUC meets three times a year to consider new and revised CPT codes and potentially misvalued services that were identified either through its Relativity Assessment Workgroup or by CMS. The RUC is supported by an Advisory Committee of 123 specialty societies that collect data and make recommendations on the work RVUs and physician time (as well as practice expenses) for the codes that the RUC has referred to them.[7] The RUC reviews the specialty society recommendations and makes recommendations to CMS (AMA, 2014b).

The predominate mechanism to establish new or revised physician work RVUs is physician surveys conducted by the specialty societies. The typical survey for a surgical procedure requests that the respondent complete the following steps:[8]

1. Select a reference code that is most similar to the survey code and the typical patient described in the survey vignette.
2. Estimate how much time is required for pre-service evaluation on the day before the procedure (90-day global only) and for pre-service evaluation time, positioning time, scrub time, intra-service time, and immediate post-service time on the day of the procedure.
3. Estimate how many post-operative visits by type are typically performed during the global period.
4. Rate the pre-service, intra-service, and post-service complexity of both the survey code and the reference code on a scale of 1 to 5 and compare the two procedures on different domains of intensity (mental effort and judgment, technical skill/physical effort, and psychological stress) using a scale of 1 to 5.

---

[7] The Advisory Committee is comprised of a physician representative appointed by each of the specialties seated in the AMA House of Delegates. Other specialties may be invited to participate in valuing services that are of particular relevance to their members.

[8] The survey also obtains demographic information on specialty, geographic location, primary type of practice, and disclosure of any direct financial interest in the procedure (other than providing the service). For the survey code, the survey asks whether the respondent's typical patient is similar to the patient described in the vignette, where the service is typically performed, and how often the procedure is performed. For a hospital procedure, the respondent is asked whether the patient is kept overnight and, if so, whether an evaluation and management (E&M) service is performed on the day of the procedure.

5. Estimate total work RVUs for the survey code using magnitude estimation and the work value for the reference code. Separate estimates for the individual work components are not obtained.

The RUC uses a variety of methodologies to review the specialty-society proposals and formulate its recommendations, including a comparison of the proposed values to values for key reference codes furnished by the specialty and by other physician specialties using the Multi-Specialty Points of Comparison (MPC) listing. An objective in these comparisons is to maintain relativity within families of codes and across different services. This process of maintaining relativity is described in greater depth in Chapter 10.

CMS reviews the RUC recommendations on a code-by-code basis. The agency's clinical review generally includes review of the methodology and rationale provided by the RUC, medical literature, and comments from the Medicare Payment Advisory Commission and the public received during the rulemaking process; consultation with CMS and other government physicians; and comparison with the RVUs for other codes. The review also draws on the clinical experience of the physicians on the review team (CMS, 2011).

The update and review process has evolved over time.[9] Beginning with the second five-year review, the building block method (BBM) has been used as an analytic tool to identify potentially misvalued surgical codes. It assumes that the total work valuation comprises building blocks composed of pre-service, intra-service, and post-service work components.

$$\text{Total RVU} = \text{pre-service work} + \text{intra-service work} + \text{immediate post-service work} \\ + \text{post-operative E+M visit work}$$

Using this formula, one can also use a "reverse" BBM to derive values for each component by subtracting the values for the other components from total work. In particular, the reverse BBM has been used to derive estimates of intra-service work and intensity, which is defined as intra-service work per unit of time (IWPUT).

9 Because of the timing of the CPT Editorial Panel decisions (which lead to the need to value new or revised codes), the RUC recommendations, and the rulemaking process, CMS has been publishing its decisions on the work RVUs on new or revised codes as interim final values subject to public comment each November and proposing final values for these codes, as well as revised values for potentially misvalued codes, each May. If public comments substantiate that refinement to the interim work values should be considered or CMS desires more clinical input on the RUC's recommendations, CMS refers the code to a multispecialty refinement panel for assistance in determining the final work RVUs. The results of the review and response to the comments are provided in the succeeding year final rule when CMS adopts the final work RVUs for the code (CMS, 2011). In the final rule for the 2015 physician fee schedule, CMS announced that beginning with 2017, the agency will propose values for most new, revised, and potentially misvalued codes and will consider public comments before establishing final values. Interim final values will be utilized only for new services for which there are no prior codes and for which RUC recommendations are not received in time for CMS to propose values. CMS is evaluating whether this change eliminates the need for a refinement panel (CMS, 2014b).

$$\text{Intra-service work} = \text{Total work} - \text{pre-service work} - \text{immediate post-service work} - \text{post-operative E+M visits}$$

$$IWPUT = \frac{\textit{Intra-service work}}{\textit{Intra-service time}}$$

The IWPUT values have been used to align the values for procedures within a family of codes and to validate the RVUs for a given procedure. However, there are concerns regarding whether the derived IWPUT values for short-duration procedures of high intensity reflect the intensity of those procedures and whether IWPUT should be used to validate RVUs for cognitive services (CMS, 2009).

Beginning in 2008, CMS and the RUC began to review codes that were potentially misvalued on an annual basis. The screening criteria have included codes where there has been a shift in site of service (i.e., codes that were valued as inpatient services that are now provided as outpatient services), high-volume growth, high IWPUT, and services that were surveyed by one specialty and are now performed by a different specialty. The ACA added a requirement that potentially misvalued services in seven categories be examined as appropriate: high-growth codes, codes with substantial changes in practice expenses, codes for new technologies or services, multiple codes frequently billed in conjunction with furnishing a single service, low-value codes often billed multiple times during a single encounter, other codes identified as potentially misvalued by the secretary (such as codes with low work values but high volume and codes on the MPC), and codes that had not been reviewed since their initial valuation in the Harvard study (CMS, 2014b).

Revisions in the work RVUs are subject to a budget neutrality constraint. If the changes are expected to cause total physician fee schedule payments to differ by more than $20 million from what the payments would have been in the absence of the RVU revisions, an adjustment is made. When CPT coding changes are made, the budget neutrality adjustment may be made to the work RVUs for the family of affected codes so that structural coding changes do not affect the aggregate work values for the services unaffected by the coding changes. Otherwise, an adjustment is made to the conversion factor to preserve budget neutrality.

Concerns with the Current System

It is important that work RVUs be accurately set under the physician fee schedule to ensure access to medically appropriate services. If a procedure is overvalued, an incentive may be created to provide unnecessary services; if a procedure is undervalued, the service may be hard to obtain and potential access problems could result. Moreover, systematic over- or undervaluing of procedures furnished by particular specialties can distort overall physician compensation levels and affect the specialty choices made by new physicians.

Below, we summarize several methodological concerns that have been raised with the current system.

The RUC Process Is Potentially Biased

MedPAC and others have raised continuing concerns about the advisability of relying on the recommendations of specialty societies that have a vested interest in the outcome of review process. In its March 2006 report, MedPAC recommended that CMS seek advice independent of the RUC, indicating that CMS needs a regular source of expertise to assist in the highly complex process of valuing practitioner services (MedPAC, 2006). These concerns led to the inclusion in the ACA requirements that CMS establish a formal process to validate RVUs that uses other data sources outside of the RUC's surveys.

The RUC Depends on Physician Surveys

The RUC relies on physician surveys that use magnitude estimation to judge subjectively the relativity of work (MedPAC, 2011a). Related concerns are the potentially small number of respondents to the surveys, low response rate, and unclear generalizability of the responses. These concerns are particularly important given the size and scope of available external data sources that have thousands of observations for the time component. In this regard, MedPAC has recommended that CMS collect current, objective time data to establish more accurate relative values (MedPAC, 2011a).

Undervalued Services Have Been Disproportionately Reviewed

The Medicare law requires that CMS review the RBRVS at least every five years. Historically, reviews of codes under the five-year review process were commonly triggered by requests from specialty societies for what they perceived as undervalued services, and the RUC's recommendations have generally been to increase values applied to undervalued services (Weems, 2009; Weems, 2008; AMA, 2012). There is no reason to expect that more procedures would become undervalued rather than overvalued over time. If anything, one would expect the opposite, given the changes in patterns of care—shorter lengths of stay, shift of services to outpatient settings, and increased proficiency with new technologies.

Beginning in 2008, CMS and the RUC began to review a number of codes that were potentially misvalued on an annual basis. The review of codes identified as potentially misvalued procedures through systematic screens employed by the RUC's Relativity Assessment

Workgroup or by CMS has resulted in more reductions for overpriced procedures than increases for underpriced procedures.[10]

Procedure Times Are Too High

Through specialty society surveys conducted as part of the RUC review process, estimates on the pre-, post-, and intra-service time required to perform services are collected. Studies have shown that RUC physician-estimated times exceed the time required to provide the services seen in external data sources (McCall, Cromwell, and Braun, 2006; Rich, 2007; Cromwell et al., 2010).

Derived Intensity Values Are Sometimes Nonsensical

The intensity component of physician work is more problematic than time to validate because it is indirectly estimated (work divided by time) and it is not well-defined. Based on findings that survey respondents confounded time and intensity, the Harvard study used magnitude estimation, whereby respondents assessed *intra-service work* instead of valuing time and intensity separately. These ratings of work were found to meet tests for internal and external validity. The current RUC survey process asks respondents to estimate *total work* (as opposed to intra-service work) using magnitude estimation. We are not aware of studies that have attempted to validate the derived intensity values from the current surveys. However, there are clear problems with face validity for some procedures. Our analysis of the 2014 physician work RVUs using the reverse BBM shows that about 3 percent of intra-service intensity (IWPUT) values are negative.[11]

CMS May Be Overpaying for Post-Procedure Care in the Global Period

For many surgical services, physicians are paid for related care within a global period (0, 10, or 90 days). The Office of the Inspector General conducted independent chart reviews of a sample of procedures and found that the number of post-procedure visits actually provided was often fewer than what the RUC had estimated for this care (Department of Health and Human Services, 2012a, 2012b). Others found increased billings by hospitalists, intensivists, and

[10] A progress report from the RUC's Relativity Assessment Workgroup indicates that of 1,543 codes that have been reviewed as potentially misvalued, the RUC recommended that work be increased for 131 codes, decreased for 567 codes, and maintained for 477 codes. There were also 250 code deletions from CPT (AMA, 2014a). In contrast, for the fourth five-year review (which was implemented in the 2012 fee schedule), the RUC recommended that the work values be increased for 83 codes, maintained for 144 codes, and decreased for 41 codes. The RUC referred 52 codes to the CPT Editorial Panel to consider coding changes prior to evaluating the work value (AMA, 2012). In 2012, CMS consolidated the five-year review and the annual review of particular codes of concern (e.g., new or revised CPT codes and potentially misvalued codes) into a single annual review process.

[11] This applies only to codes with defined IWPUT values. In some cases IWPUT values are undefined because the denominator (time) is zero in the CMS estimates. The 5-percent statistic is for the procedures on which we focus in this report. We define these codes later in the report.

nonphysician practitioners for post-operative care being provided during the global period (Cromwell et al., 2010). This raises concerns about duplicate payment for at least some of these post-operative visits. In response to these concerns, CMS announced in the final rule for the 2015 physician fee schedule that 10-day and 90-day global periods would be phased out beginning in 2017 (CMS, 2014b).

RUC Process Does Not Adequately Address Efficiency Gains

MedPAC has described how efficiency gain is common in health care (MedPAC, 2006). Physician learning by doing, technology diffusion and substitution, and changes in practitioner mix all lead to a situation in which the work and time related to services are typically decreasing. However, infrequent updating of work RVUs for individual services may not sufficiently account for these efficiency gains.

The RUC Process May Contribute to the Underpayment of Primary Care

Forecasted shortages in primary care physicians and the gap in compensation between primary care physicians and specialists have raised concerns that the RUC process is biased against primary care physicians and that primary care services are undervalued relative to procedural services (Bodenheimer, Matin, and Laing, 2008; Sinsky and Dugdale, 2013; Berenson, 2010).[12] Responding to the access concerns, the ACA provided for a temporary 10 percent Medicare incentive payment to eligible physicians and nonphysician practitioners for specified primary care services through 2015.

Recent Changes in RUC Process

Recently, the RUC introduced changes to address some of the concerns expressed with its valuation process. In 2012, the RUC increased primary care representation on the committee by adding a permanent seat for the American Geriatrics Society and a rotating seat for an actively practicing primary care physician. Several changes are directed at increasing the transparency of the process, such as publishing the RUC meeting dates, minutes of each RUC meeting, and the vote total for each individual CPT code on the RUC website. Other changes involve the development of a centralized online survey process with survey and reporting improvements and an increase in the minimum number of survey responses required for commonly performed codes (AMA, 2013).

[12] A recent study analyzing differences in intensity per unit of time reaches a different conclusion. Using the procedure times in the CMS files, the study did not find evidence of systematic higher values for physician work in procedure/test codes than in E&M codes (Berber et al., 2014). However, the study results are predicated on an assumption that there are no systematic differences in the CMS times across types of services. Other studies suggest that procedure times for surgical services may be disproportionately overstated relative to E&M services (McCall et al., 2006; Cromwell et al., 2006).

Project Objectives

RAND's objective is to develop a model for independent estimates of total work RVUs and its components: intra-service time, intra-service intensity, pre- and immediate post-service work, and post-operative visits that are included in the global period. The independent estimates provide a mechanism for validating the work RVUs. We use independent data sources to measure the characteristics of individual procedures (e.g., place of service) and characteristics of patients who receive these procedures. As discussed in greater detail in Chapter 3, we focus our investigation on the broad CPT category of surgical services. This category includes surgery (where there is an incision of some kind) and other types of procedures, such as colonoscopy, that do not require an incision. We exclude surgical procedures rarely performed in the Medicare population or almost always performed in office settings. To this list we add selected medical procedures, such as intervention cardiology services, that are similar in nature and often performed in an operating room setting. We chose this set of 3,179 procedure codes primarily because of the availability of independent data that could be used to estimate intra-service times for these services. We refer to this set of services as the "core procedures" for our modeling effort. We use the core procedures as an illustration of how the work RVUs might be predicted using independent data sources.

Our general approach for modeling RVUs for the core procedures has three components:

1. *Deriving current RVUs for the individual work components*. We use the current work RVUs and supporting documentation from Medicare physician fee schedule time files to derive initial RVUs for each work component using the reverse BBM.
2. *Estimating intra-service time and other characteristics for procedures.* We use independent data sources to estimate intra-service time for each procedure (which we call RAND time estimates) and other characteristics, such as mortality rate and years of training of physicians who perform the procedure. Intra-service time is particularly critical because it is the key driver for work RVUs.
3. *Predicting work RVUs for the individual work components and total work.* We use regression analysis to predict values for each work component and total work. Our prediction models consider the initial RVUs that we derived for the individual work components of each procedure, the RAND time estimate for the procedure, and other characteristics of the procedures and the patients on whom the procedures are performed.

The validation process we use is a complex one, and we are cognizant that there is no single optimal approach. Each of the three components described above can be done in several ways, and we examine the impact of different approaches to addressing specific methodological issues and data limitations. We compare our predicted RVUs to the current values for each work component and for total work. Because our models start with the current work RVUs, the predicted values for the core procedures are scaled in RVUs and maintain relativity across the core procedures. However, this also means that, to the extent that there are systematic biases in the current work RVUs, these are built into our predicted values. Later in the report, we discuss how changes in the average work RVUs for the core procedures might impact relativity with non-core procedures.

We believe that the RAND model estimates for the individual work components and the overall work RVUs could be used for two key applications: (1) CMS could flag codes as potentially misvalued if the CMS and RAND model estimates were notably different, and (2) CMS could also use the RAND estimates as an independent estimate of the work RVUs to consider when assessing a RUC recommendation.

As noted earlier, this project focuses on surgical procedures and selected medical procedures typically performed in an operating room setting for which independent time data are available. While the project focuses on these procedures, we also discuss the extent to which some aspects of our model could be generalizable to other types of services.

Organization of This Report

The remainder of this report is organized as follows:

- Chapter 2 starts with an explanation of the BBM. We then provide descriptive statistics of current total work and the components of work derived from current CMS estimates.
- Chapter 3 provides an overview of our modeling approach and the data that we use. We describe how we selected the services that we focus on in this project and our method of assigning those services into clinically coherent groupings. We also describe variables that are used in our prediction models. Additional details are reported in the relevant chapters that follow for a specific work component.
- Chapter 4 discusses how we estimate intra-service times and reports our estimates of intra-service times for the typical setting and across all settings in which a service is provided.
- Chapters 5 through 7 pertain to our modeling efforts for each of the individual work components: preoperative and immediate post-operative work (Chapter 5), post-operative visits (Chapter 6), and intra-service work and intensity (Chapter 7). Each chapter discusses the assumptions that are incorporated into alternative prediction models and presents the values from our prediction models for the individual component.
- Chapter 8 consolidates the results from our individual prediction models in Chapters 5 through 7 into our estimates of total work using alternative approaches. It compares these results using a BBM approach to a model that uses a single prediction model to directly estimate total work values. It explores the differences between the models and summarizes the percentage change from current work RVUs by procedure characteristics and by specialty.
- Chapter 9 considers selected topics that are not dealt with in our model. We discuss our analyses of the incremental time associated with performing multiple procedures and add-on procedures to base surgical codes and the potential impact of observation services

on our results.[13] We also discuss two issues raised by CMS in its 2015 rulemaking for the physician fee schedule: valuing conscious sedation administered by the physician performing the procedure and implementing 0-day global periods for surgical procedures currently paid under 10-day or 90-day global periods.

- Chapter 10 summarizes our findings and the limitations of our prediction models. It discusses the implications of our findings and next steps in developing a validation model for total work RVUs. We also address questions of maintaining relativity and generalizability.

We have also provided additional information on our modeling approach and results in appendixes. Appendixes A–C describe how we selected the core procedure codes for our analyses, explain the code groupings that we use in our predictive models, and explain the variables that we use in our predictive models. Appendixes D and E provide detailed explanations of the methods we used to transform anesthesia and operating room times into surgical times and to pool data to estimate surgical times. Details on analyses looking at the incremental impact of multiple and add-on procedures on surgical times are reported in Appendix F. Appendixes G and H contain the output from our prediction models and additional analyses of the differences between the CMS and RAND estimates.

[13] These are the services provided to a Medicare beneficiary who remains under observation as an outpatient instead of being admitted as an inpatient.

2. Descriptive Statistics of Total Work and Components of Work

Overview

In this chapter we provide a descriptive analysis of how time and work are distributed for the core study procedures in current CMS estimates across each component of the BBM: intra-service work, pre- and immediate post-service work, and, for surgical services with global periods, related post-operative E&M visits. This chapter begins with a description of the CMS estimates and the BBM. We then review a number of descriptive analyses. These analyses set the stage for understanding the relative importance of the individual work components in the current system and the analyses that we undertake in the remainder of the report to validate those components under our proposed validation models.

CMS Estimates and Work RVUs

As part of its annual rulemaking process, CMS produces a physician time file that lists time estimates (in minutes) for many, but not all, work components of each service, including intra-service time, pre-service time, and immediate post-service time. For procedures with global periods, the file also includes counts of the post-operative E&M visits that are bundled into the physician work RVUs. The physician time file data reflect CMS review and adjustment of initial time and E&M visit counts from the specialty society surveys and RUC recommendations.

Addendum B to the Medicare physician fee schedule includes work RVUs for each service. Like the estimates in the physician time file, these work RVUs typically originate from the RUC's recommendations and are reviewed and adjusted by CMS.

We refer to the time estimates from the time file and work RVUs from Addendum B of the Medicare physician fee schedule as "CMS estimates" because they are subject to CMS review and revision.

Overview of the Building Block Method

The BBM assumes that total work RVUs are equal to the sum of the RVUs associated with each of the components in the BBM, including the pre-service, intra-service, immediate post-service, and post-operative E&M components. These RVUs are the product of the time associated with each BBM component measured in minutes and intensity measured in RVUs per minute.[14]

[14] Not all procedures include each of the components of the BBM. For example, only procedures with a global period have post-operative visits. The fraction of procedures that has a given component is described in Table 2.2.

$$\begin{aligned} Total\ Work\ RVUs \\ &= (\textit{Pre-service minutes} \times \textit{Pre-service intensity}) \\ &+ (\textit{Intra-service minutes} \times \textit{Intra-service intensity}) \\ &+ (\textit{Immediate post-service minutes} \times \textit{Immediate post-service intensity}) \\ &+ (\textit{Post-operative visits} \times \textit{RVUs per post-operative visit}) \end{aligned}$$

Some components are broken down into even more detailed categories. For example, the CMS time file lists three subcomponents for pre-service time: pre-service evaluation, pre-service positioning, and pre-service scrubbing. The same two rules apply under the BBM regardless of the level of detail: First, total work is the sum of work associated with individual components; and second, the work associated with an individual component is the product of time and intensity.

The work associated with post-operative E&M visits follows these rules, with a slight difference. The CMS time file records a number of visits by E&M code rather than minutes. Still, because each E&M visit code is itself assigned a number of minutes and a total number of work RVUs, it is possible to calculate the time and intensity for each E&M visit.

Table 2.1 below illustrates the BBM approach to calculating total work for CPT 47562 (laparoscopic cholecystectomy), CPT 43239 (esophagogastroduodenoscopy [EGD] with biopsy, single or multiple), and CPT 33322 (suture repair of aorta or great vessels; with cardiopulmonary bypass).

The estimates for each work component come either directly from the CMS estimates or are calculated using those estimates. Estimates of total work, intra-service time, pre- and immediate post-service time, and the number of post-operative visits come directly from the CMS estimates. The amount of work tied to each type of post-operative visit comes from how much work is assigned to these visits in the current RBRVS system. We assume constant intensity for the time spent in pre- and post-service components based on the intensity used by the initial Harvard estimates. Using these estimates, we can calculate the values for the other components. For example, pre-service work comes from multiplying pre-service time and pre-service intensity. We have only intra-service time from the CMS time file. Intra-service work is derived by subtracting the work values for the other components from total work RVUs.

$$\begin{aligned} Intra\text{-}service\ work \\ &= Total\ work - pre\text{-}service\ work - immediate\ post\text{-}service\ work \\ &- post\text{-}operative\ visit\ work \end{aligned}$$

Having an estimate for both intra-service work and intra-service time, we can then derive intra-service intensity (IWPUT).

$$IWPUT = \frac{Intra\text{-}service\ work}{Intra\text{-}service\ time}$$

The breakdown of the work RVUs for the laparoscopic cholecystectomy (Table 2.1.A) illustrates the many different components that can go into a complex procedure with a 90-day global period. For this procedure, work RVUs associated with pre- and post-operative visits account for nearly the same amount of work for the actual operation (intra-service work). This breakdown of pre- and post-service work versus intra-service work is discussed in more depth later in this chapter. The EGD procedure has fewer components, as there are no post-operative visits. With the EGD, the performance of the procedure (intra-service work) makes up the majority of RVUs (1.67 out of 2.47) but only about a quarter of the time (15 minutes out of a total of 54 minutes). The aortic repair procedure illustrates a problematic aspect of the BBM—namely, that intra-service work and IWPUT can be negative when the work associated with the pre-service, immediate post-service, and post-operative E&M components are themselves larger than total work. The total work estimated for this procedure is 24.42 work RVUs. There are 14.5 post-operative visits after the operation in the 90-day global period. The work RVUs associated with these 14.5 post-operative visits total 29.49. This is obviously nonsensical because it exceeds the total work RVUs (24.42) and leads to an estimate of *negative* intra-service work and negative IWPUT using the BBM. The negative values are an indication that the code is potentially misvalued and should be reviewed.

Table 2.1. Work RVU Components for Selected Surgical Procedures from CMS Estimates

A. CPT 47562 (laparoscopic cholecystectomy)

Component			Minutes	Intensity	Work (RVUs)
Immediate pre-service			**65**	**0.019**	**1.24**
Evaluation			*40*	*0.022*	*0.90*
Positioning			*10*	*0.022*	*0.22*
Scrub			*15*	*0.008*	*0.12*
Intra-service			**80**	**0.074**	**5.94**
Immediate post-service			**25**	**0.022**	**0.56**
Post-operative E&M			**58**	**0.047**	**2.73**
	Number	Mins/ea.			
99212 (outpatient visit)	*1*	*16*	*16*	*0.030*	*0.48*
99213	*1*	*23*	*23*	*0.042*	*0.97*
99238 (discharge)	*0.5*	*38*	*19*	*0.067*	*1.28*
TOTAL			**228**	**0.049**	**10.47**

B. CPT 43239 (EGD diagnostic, single or multiple)

Component			Minutes	Intensity	Work (RVUs)
Immediate pre-service			**27**	**0.020**	**0.53**
Evaluation			*19*	*0.0224*	0.43
Positioning			*3*	*0.022*	0.07
Scrub			*5*	*0.008*	0.04
Intra-service			**15**	0.111	**1.67**
Immediate post-service			**12**	**0.022**	**0.27**
Post-operative			**0**	**-**	**–0**
	Number	Mins/ea.			
99291 (critical care E&M)	*0*	*0*	*0*	*0*	*0*
99231 (inpatient visit)	*0*	*0*	*0*	*0*	*0*
99232 (inpatient visit)	*0*	*0*	*0*	*0*	*0*
99233 (inpatient visit)	*0*	*0*	*0*	*0*	*0*
99212 (outpatient visit)	*0*	*0*	*0*	*0*	*0*
99214 (outpatient visit)	*0*	*0*	*0*	*0*	*0*
99238 (discharge)	*0*	*0*	*0*	*0*	*0*
TOTAL			**54**	**0.046**	**2.47**

C: CPT 33322 (suture repair of aorta)

Component	Minutes	Intensity	Work
Immediate pre-service		**0.019**	**1.79**
	96		
Evaluation	*41*	*0.022*	*0.92*

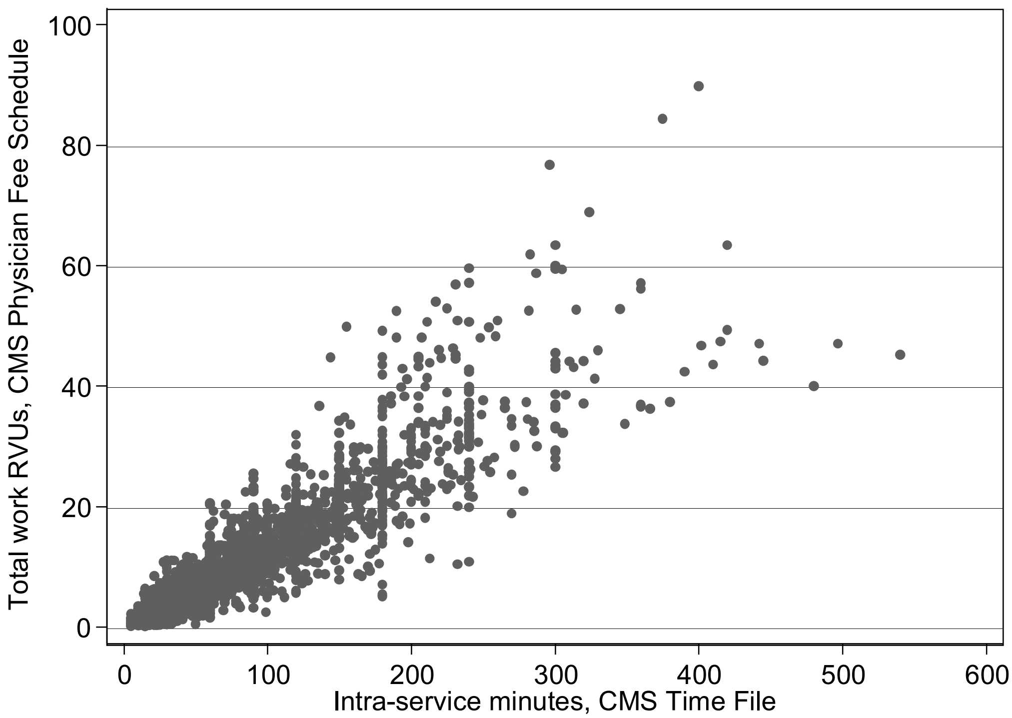
Component			Minutes	Intensity	Work
Positioning			*30*	*0.022*	*0.67*
Scrub			*25*	*0.008*	*0.20*
Intra-service			**175**	**–0.046**	**–7.98**
Immediate post-service			**50**	**0.022**	**1.12**
Post-operative	Number	Mins/ea.	**563**	**0.053**	**29.49**
99214 (outpatient visit)	*1.5*	*40*	*60*	*0.038*	*2.25*
99231 (inpatient visit)	*7.5*	*20*	*150*	*0.038*	*5.70*
99238 (discharge)	*1*	*38*	*38*	*0.034*	*1.28*
99291 (critical care E&M)	*4.5*	*70*	*315*	*0.064*	*20.25*
TOTAL			**884**	**0.028**	**24.42**

NOTES: Bolded rows are added together to calculate the total minutes and RVUs reported in the bottom row. All time estimates, post-operative visit counts, and total work are from the CMS estimates. Post-operative E&M visit work and intensity are imputed from the work RVUs and times calculated separately for those services. Pre- and immediate post-service components have assumed intensity values. Other intensity values are calculated by dividing work by minutes. Intra-service work and intensity are derived after the work values for the other work components are calculated.

Total Work and Time

Total work and intra-service time are strongly positively correlated in a "core group" of procedures included in our analyses (Figure 2.1). The core group of procedures is described in the next chapter. The correlation coefficient between intra-service time and total work in the CMS estimates is 0.91. Each additional minute of intra-service time is associated with an additional 0.14 RVUs on average. The relationship between work and time can also be expressed as an elasticity: A 10-percent increase in CMS-reported total time is associated with a 4.9-percent increase in CMS-reported total work. This strong correlation emphasizes that the most important procedure characteristic in determining total work RVUs is the intra-service time.

Figure 2.1. CMS-Reported Time Versus Work



Distribution of Work and Time Components

Under the BBM, total work is equal to the sum of pre-service, intra-service, immediate post-service, and post-operative visit work. In the same way, total time is the sum of the times associated with individual components. We use the CMS estimates and the reverse BBM to decompose the total work and time associated with each core procedure into the four work components. Figure 2.2 illustrates the results of our decomposition of total work and time and highlights the relative importance of each component. We find that intra-service work accounts on average for 46 percent of total work RVUs compared to 33 percent of total minutes. The second most important component is post-operative E&M visits, which account for one-third of both total time and total work. Pre-service work accounts for 22 percent of total minutes but only 13 percent of total work, while immediate post-operative work accounts for 12 percent of time but 8 percent of total work.

Figure 2.2. Average Contribution of Individual Components to Total Work and Time for Core Procedures

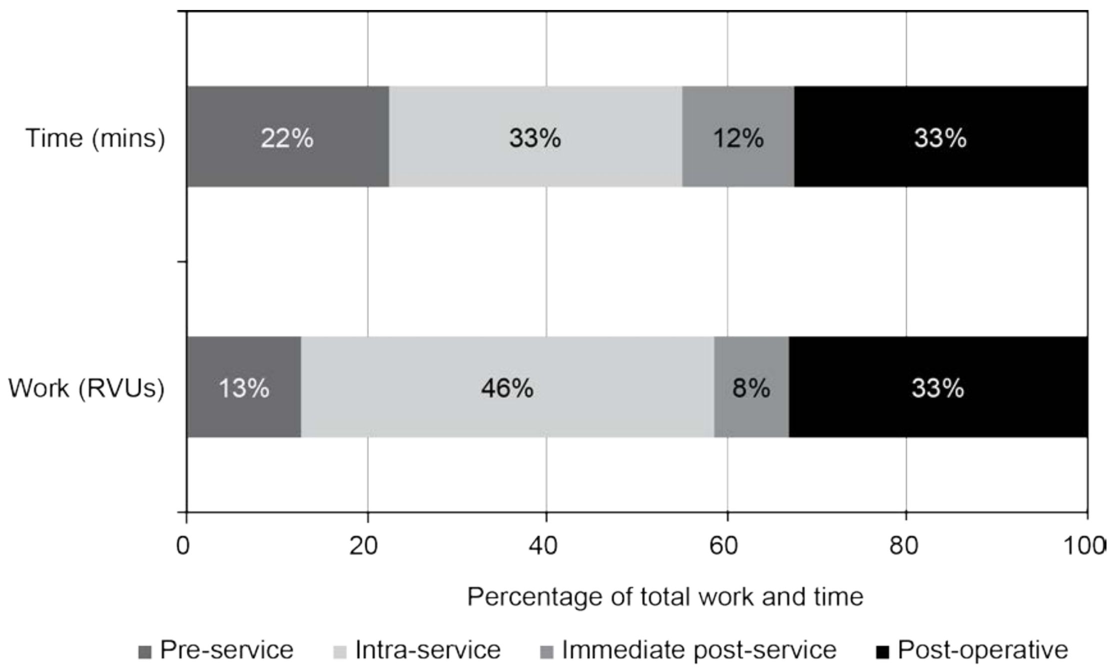


Table 2.2 reports summary statistics for each work category for only those core procedures with nonzero reported work for a given component. The largest two components of work are the intra-service component and the post-operative E&M visits component. About 20 percent of core procedures have a 0-day global period (in which case only visits occurring on the day of the procedure are bundled) or are not subject to the global period policy (in which case any visits related to the surgical procedure are separately payable).

Table 2.2. Descriptive Statistics for Work (RVUs) Across Work Components for Core Procedures Average Across Only Those Codes with Nonzero Value for That Component

Work Component	% of CPT Procedures with Work>0 RVUs	Mean Work RVUs	St. Dev.	25th Percentile	Median	75th Percentile
Total	100	10.65	10.34	3.45	7.39	14.59
Pre-service	98.8	0.91	0.59	0.49	0.78	1.28
Intra-service	100	4.83	5.71	1.28	2.98	6.29
Immediate post-service	98.7	0.58	0.66	0.34	0.45	0.67
Post-operative E&M visits	79.6	5.47	5.41	2.08	3.88	7.23

Figure 2.3 illustrates the separate distributions of work values associated with each component of physician work. The figure highlights that when the reverse BBM is used to derive

intra-service work, approximately 3 percent of the core procedures have negative intra-service work values. This illustrates one concern discussed in Chapter 1 with the current process. Figure 2.3 also reinforces two earlier observations: (1) The intra-service and post-operative E&M visit components account for the majority of work, and (2) the distributions of individual components are right-skewed—i.e., there are outlier surgical codes with large work values compared to other codes.

Figure 2.3. Distributions of Work RVUs for Core Procedure Codes by Individual Component[15]

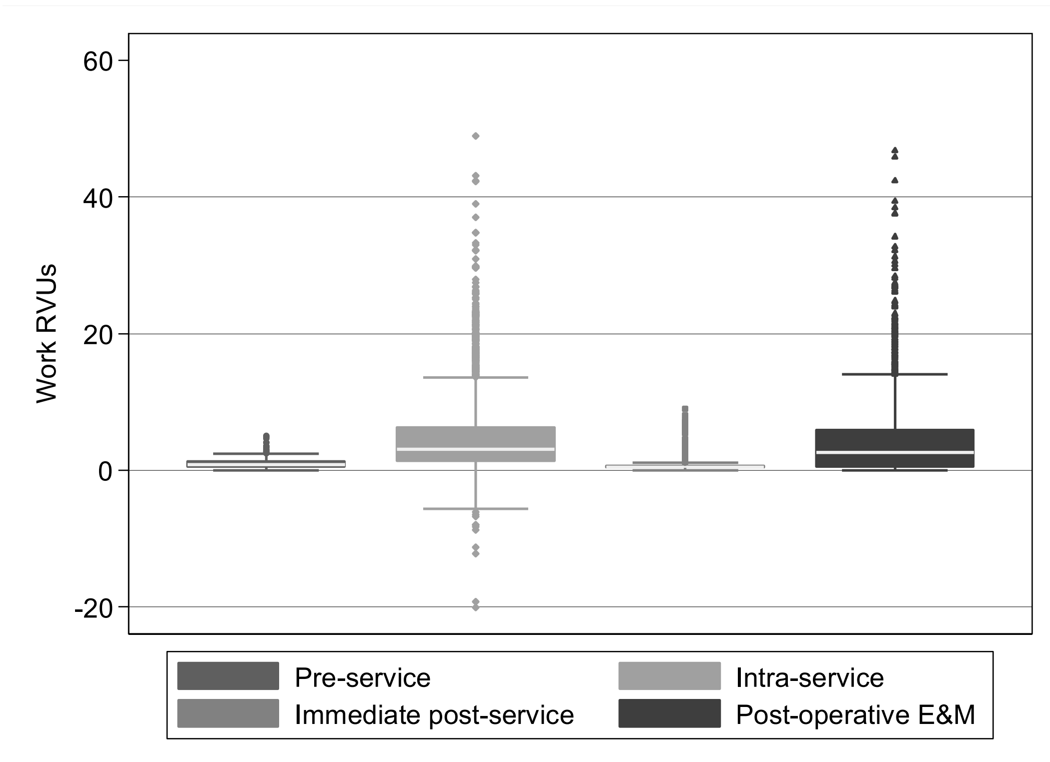


Figure 2.4 breaks down total work into the individual components by the global period assigned to the code. The average RVUS for post-operative E&M visits are, as expected, largest for surgery codes with the longer, 90-day global period (69 percent of all surgery codes). The RVUs for post-operative E&M visits are a smaller share of total work for surgery codes with a 10-day global period and almost none of the work for codes with a 0-day or no global period.[16] Codes with a 90-day global period have significantly higher intra-service work, as well as

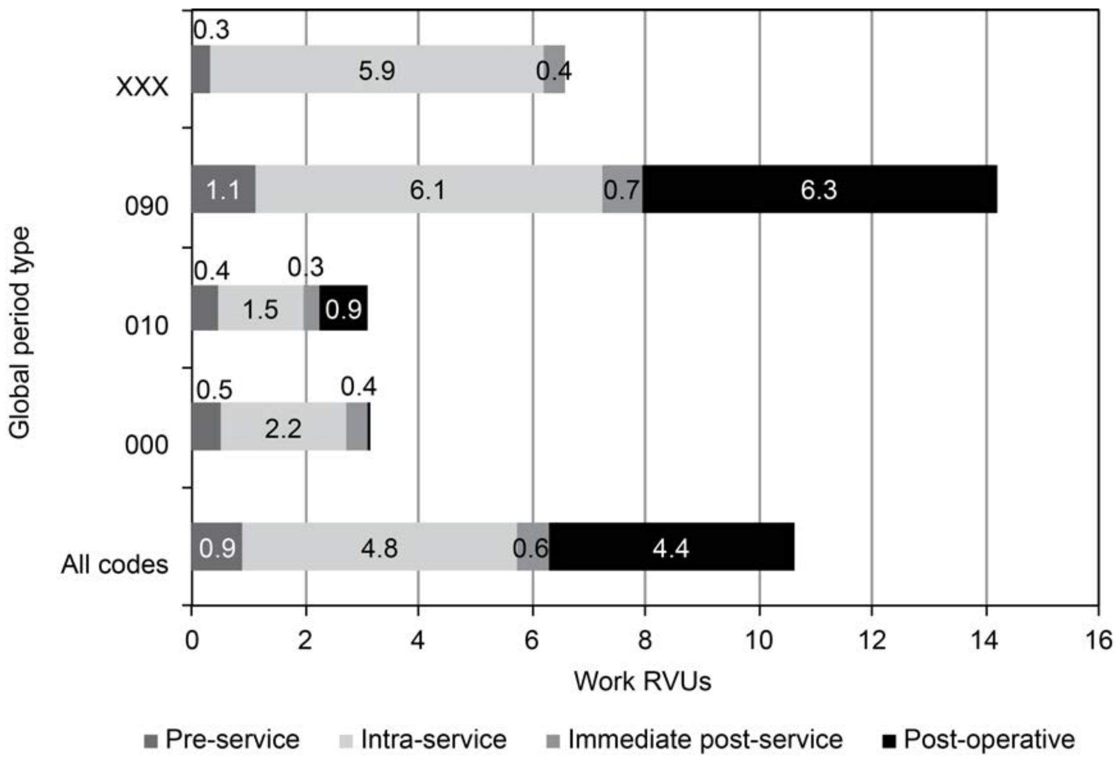
---

[15] The box plots illustrate the 25th percentile, median, and 75th percentile as the bottom, horizontal marker, and top of the solid box. The upper and lower whiskers include adjacent values, and the points illustrate outside values.

[16] We address post-operative E&M work assigned to "000" global codes in Chapter 6 and detail how we have applied corrections to address this apparent discrepancy.

significantly higher post-operative E&M visit work compared to other codes. The figure again highlights the importance of post-operative E&M visits in total work RVUs. While codes without a global period (shown as XXX in the figure) have no post-operative E&M visits work, they have on average more intra-service work (and more work overall) compared to codes with a 0-day or 10-day global period.

Figure 2.4. Average RVUs by Work Component, by Global Period[17]

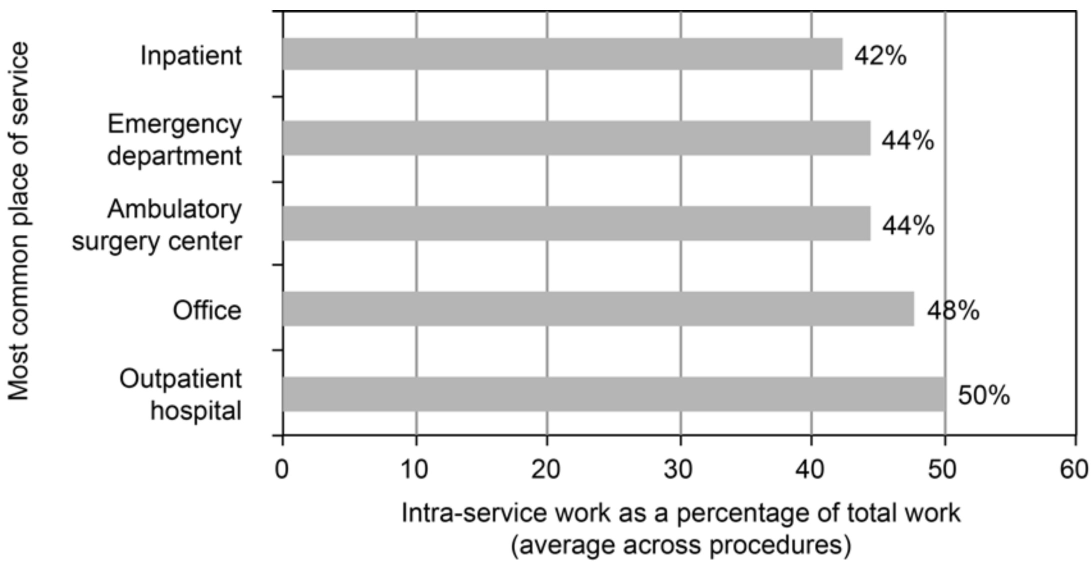


Intra-Service Work by Primary Place of Service

Intra-service work accounts for roughly the same proportion of total work across different sites of care (Figure 2.5). There are some differences. Intra-service work accounts for the smallest fraction of total work among surgical procedures most often performed in the inpatient setting, which have a higher proportion of 90-day global periods. However, in none of the settings does intra-service work account for more than 50 percent of total work.

[17] XXX is used when the global period policy does not apply to the code.

Figure 2.5. Average Proportion of Work Allotted to Intra-Service Work by Most Frequent Place of Service in Which the Procedure Is Performed



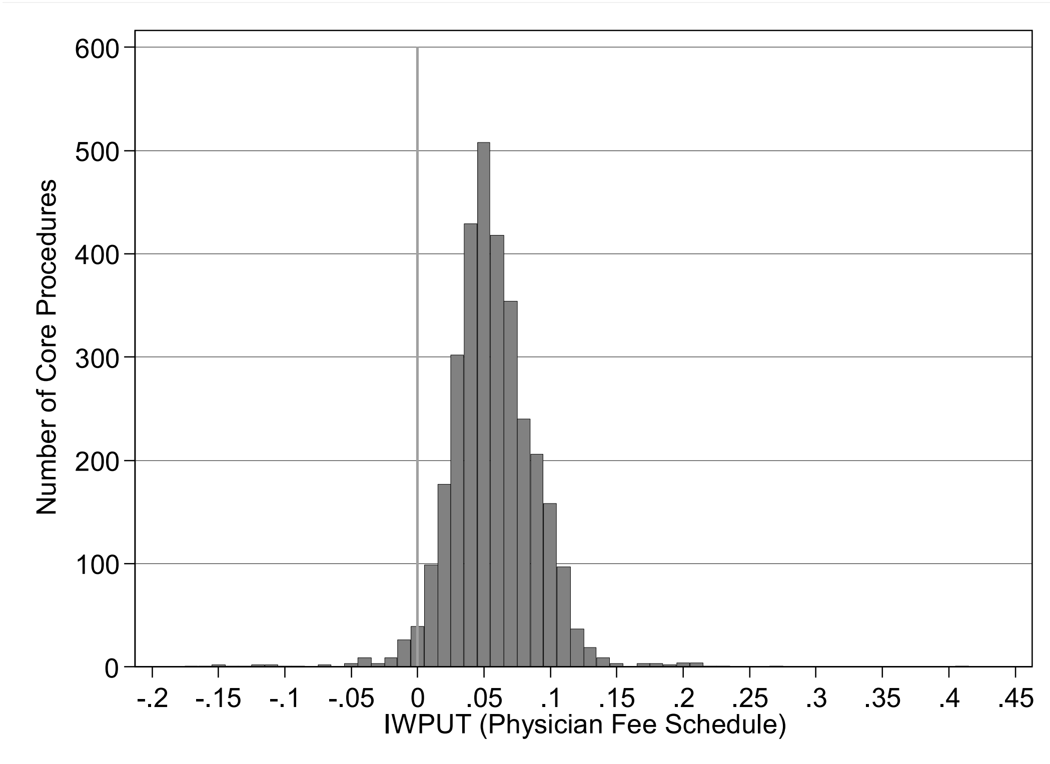
Intra-Service Intensity

As discussed earlier in this chapter, we can derive intra-service intensity per unit of work (IWPUT) from the CMS estimates using the reverse BBM method:

$$IWPUT_{derived} = \frac{\text{Total Work RVU} - \text{pre-service work}_{\text{derived}} - \text{immediate post-service work}_{\text{derived}} - \text{post-operative}_{\text{derived}}}{\text{intra-service time}}$$

In Figure 2.6, we show the distribution of IWPUT values that we derive from the CMS estimates. As points of reference, the IWPUT for most pre-service and immediate post-service work is 0.0224, and the IWPUT for CPT 32854 (lung transplant with bypass) is 0.11. We use these as reference points to create a range of intensity and to see what fraction of codes are outside that range. In the CMS time file, 15.5 percent of core procedures are outside that range, including 86 procedures (2.7 percent) with 0 or negative IWPUT, 248 procedures (7.8 percent) with IWPUT between 0 and 0.0224, and 153 procedures (4.8 percent) with IWPUT greater than 0.11.

Figure 2.6. Distribution of IWPUT Values for Core Procedures Derived from CMS Estimates



Summary

In this chapter, we summarize for the core procedures how total work and the individual work components are distributed using the CMS estimates and the reverse BBM. We find that the intra-service and E&M visit components account for the majority of work and that there is greater variation in these values than in pre-service and immediate post-service work. Not surprisingly, the contribution of the E&M visit component to total work is greatest for surgery codes with a 90-day global period. These codes (69 percent of the core procedures) also have significantly higher intra-service work compared to other core procedures. When we derive IWPUT from the CMS estimates, we find that 15.5 percent of core procedures either have an IWPUT value that is less than the value for most pre-service and immediate post-service work or greater than a CPT 32854 (lung transplant with bypass). In the chapters that follow, we examine each work component separately and use revised time estimates along with procedure and patient characteristics to predict RVUs for each component and total work. We contrast the results from these RAND models to the CMS estimates that are described in these chapters.

3. Methodological Approach and Data

Chapter Overview

We begin this chapter with a description of the core procedures that are included in our models. We then provide an overview of our modeling approach and how we explore different modeling issues. We conclude by summarizing the steps that we use to predict RVUs for total work and the work components. The details on the prediction models for each work component are provided in the relevant chapters of the report.

Procedures Included and Excluded from Our Models

As an illustration of our models, we focus on the CPT category of surgical procedures. This broad category includes surgery (where there is an incision of some kind) and other types of procedures, such as colonoscopy, that do not require an incision. To this list we add some select medical procedures, such as interventional cardiology services, that are similar in nature and often performed in operating room settings.

We choose these codes primarily because of feasibility. Our goal is to validate the different components that go into calculating work RVUs for a given service. The datasets available to us for this project have data on intra-service time for surgical services provided in hospital inpatient and outpatient settings and in ambulatory surgical centers (ASCs).

From this list of codes we exclude a number of services. These include codes that are rarely performed in the Medicare population and codes that are priced by individual carriers. This left us with a total of 3,179 codes. The details on our inclusion and exclusion criteria are included in Appendix A. The core codes account for 59 percent of codes and 85 percent of the estimated total work RVUs paid under the Medicare physician fee schedule for surgical procedures in 2012.[18]

As is typical in such analyses, a small number of codes account for the vast majority of Medicare allowed charges. The top 100 codes (3.2 percent of the 3,179 codes) account for 64.4 percent of all physician work RVUs for the core procedures (Table 3.1). This table is a useful reminder that any change in work RVUs for this small number of codes could have large shifts in physician payment. In comparing the RAND estimates from our prediction models to the CMS estimates, we highlight the changes on 20 codes that are high in volume and reflect a range of

[18] In these calculations we include medical procedures that we added to the core codes in both the numerator and denominator.

services by clinical focus and specialty. These codes are summarized in Table 3.2. Together, they account for 22 percent of the aggregate work RVUs for the core procedures in 2012.

Table 3.1. Fraction of Codes and 2012 Work RVUs Accounted by Top Codes

Top Core Codes in Terms of Total Medicare Allowed Charges	Fraction of All Core Codes	Fraction of Work RVUs for All Core Codes
Top 10 core codes	0.31%	24.41%
Top 25 core codes	0.79%	38.73%
Top 50 core codes	1.57%	50.59%
Top 100 core codes	3.15%	64.44%

Table 3.2. Overview of High-Expenditure Procedures Included in Summary Analyses of Differences Between CMS and RAND Estimates

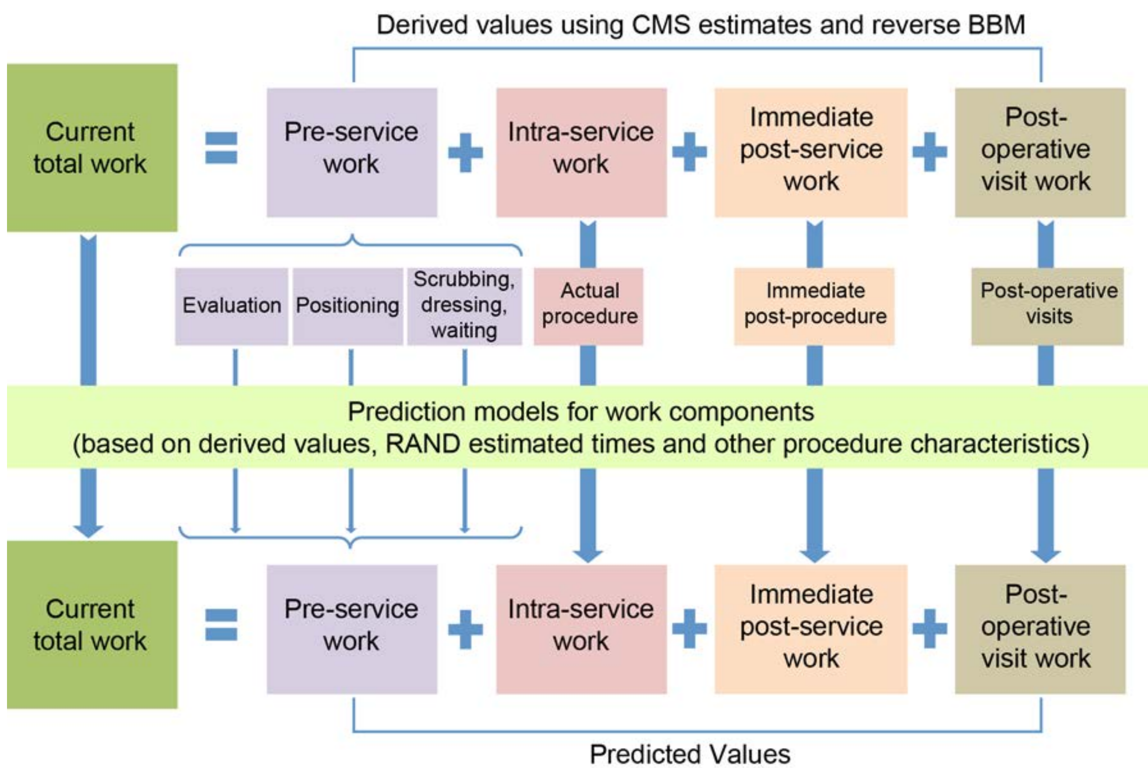
CPT	CPT Short Description	Medicare 2012 Volume (000s)	Total Work RVUs	Aggregate Work RVUs (000s)	Global Period	Typical Setting	Specialty
13132	Complex laceration repair	148.6	4.78	710.3	10	Office	ED/dermatology
17311	Mohs procedure of skin	472.1	6.2	2927.0	0	Office	Dermatology
20610	Drain or inject a bursa or joint	6095.1	0.79	4815.2	0	Office	Orthopedics
27245	Treat thigh fracture	73.5	18.18	1336.4	90	Inpatient	Orthopedics
27447	Total knee arthroplasty	262.7	20.72	5442.4	90	Inpatient	Orthopedics
31231	Diagnostic nasal endoscopy	435.5	1.1	479.1	0	Office	ENT
33533	Coronary artery bypass graft (CABG) arterial single	59.1	33.75	1995.3	90	Inpatient	CT surgery
35301	Rechanneling of artery	49.8	21.16	1054.2	90	Inpatient	Vascular surgery
43239	Esophagogastroduodenoscopy (EGD) with biopsy	1175.6	2.47	2903.8	0	ASC	Gastroenterology
44120	Removal of small intestine	24.7	20.82	514.6	90	Inpatient	General surgery
45380	Colonoscopy with biopsy	785.8	4.43	3480.9	0	ASC	Gastroenterology
47562	Laparoscopic cholecystectomy	113.6	10.47	1189.3	90	Outpatient	General surgery
52000	Cystoscopy	900.6	2.23	2008.3	0	Office	Urology
52601	Prostatectomy (TURP)	47.0	15.26	717.4	90	Outpatient	Urology
62311	Inject spine lumbar/sacral	907.1	1.17	1061.3	0	ASC	Ortho/neurosurgery
63047	Remove spinal lamina (lumbar)	68.7	15.37	1055.2	90	Inpatient	Ortho/neurosurgery
64450	Digital nerve block	719.0	0.8	539.3	0	Office	Podiatry
66984	Cataract surgery with IOL (1 stage)	1576.4	8.52	13431.0	90	ASC	Ophthalmology
67228	Treatment of retinal lesion	82.6	13.82	1142.1	90	Office	Ophthalmology
93458	Left heart artery and ventricle angiography	485.8	5.85	2842.0	0	Inpatient	Cardiology

NOTE: Discounted volumes are calculated from the utilization file for the 2014 physician fee schedule. The file crosswalks 2012 codes into their 2014 equivalents. We included only services that represent the performance of the actual surgical procedure. Volumes reported for assistants-at-surgery and pre-service or post-services only or discontinued services have not been included. The discounted volume adjusts for multiple procedure discounting, bilateral procedures, and the use of surgical teams.

Overview of Modeling Approach

Our objective is to estimate total work for a surgical procedure. We use the BBM to do this by predicting the work RVUs for each component in the BBM and, in the case of the pre-service component, the individual subcomponents (evaluation; positioning; and scrubbing, dressing, and waiting). Using the BBM, we sum the predicted RVUs for each work component to estimate total work. Alternatively, we develop a separate model that estimates total work directly (Figure 3.1). Under either alternative, three major activities are required to estimate total work: derive initial RVUs for each work component from the CMS total work values using the reverse BBM, develop measures of intra-service times from external data sources and other measures of procedure and patient characteristics for the prediction models, and predict work RVUs values using the prediction models.

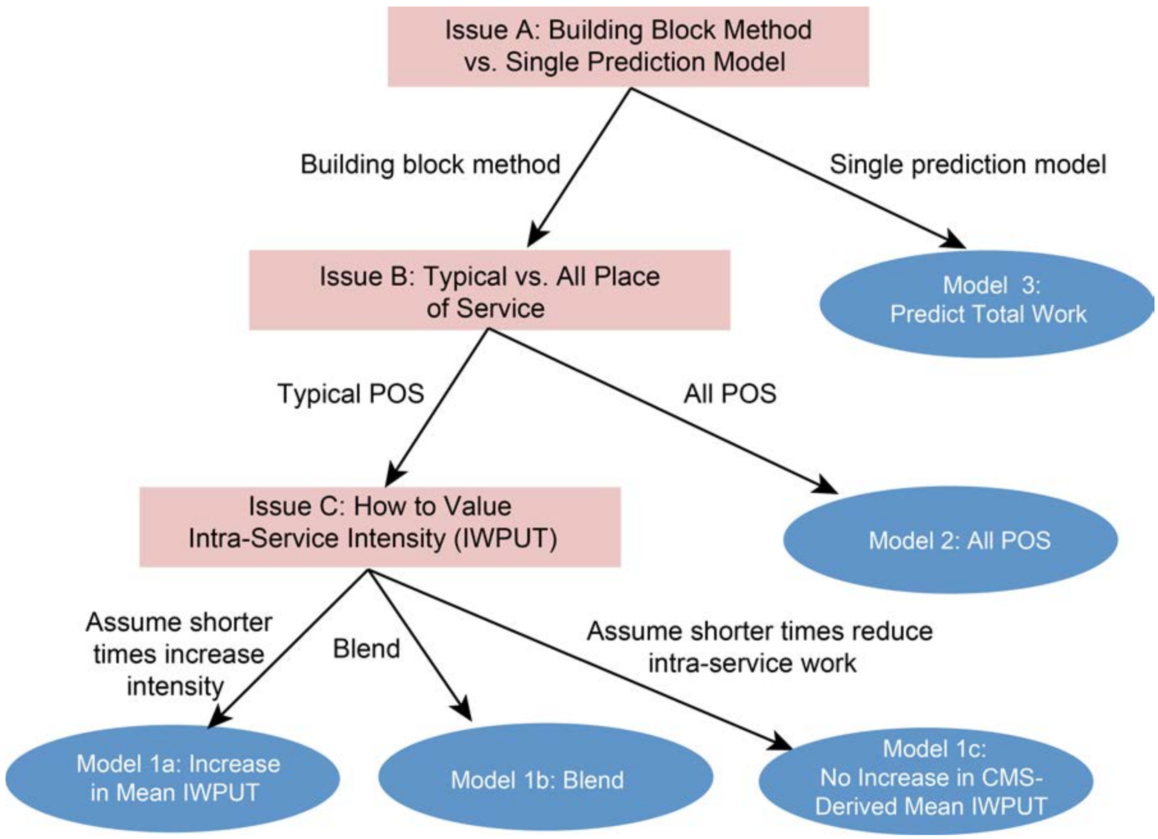
Figure 3.1. Overview of Modeling Approach



Modeling Options

The validation process is a complex one, and we are cognizant that there is no single optimal approach. In both estimating RAND times and our prediction models, there are many options we could pursue. We investigate different approaches and compare the results at each step of our model development. Figure 3.2 summarizes the three major issues that we considered in modeling total work and how we combined choices on these issues into the models that we present in this report. The models that we develop highlight the effect of these issues on total work RVUs.

Figure 3.2. Overview of Models Used to Predict Total Work



Issue A: BBM Versus Single Prediction Model to Estimate Total Work RVUs

Models 1 and 2 use the BBM, while Model 3 uses a single prediction model to estimate total work RVUs. Predicting total work in a single step is consistent with the magnitude estimation used in the RUC process to value codes. While the surveys collect information on the individual work components, the final valuation is based on the relative work involved in performing the total procedure (including post-operative visits included in a global period). For a validation model, a direct estimation of total work can indicate whether a procedure may be misvalued relative to other procedures, but a limitation is that the results do not indicate which work component drives the misvaluation. In contrast, models that use the BBM separately value the individual components so that the predicted values for each component can be compared to the values derived from the CMS estimates. Further, using the BBM approach allows for refinement of a BBM component for which external data are available without modifying the other components. However, the underlying assumption for the BBM is that accurate estimates of total work can be generated by aggregating estimates of the individual BBM components. Because the CMS total work estimates have been generated through magnitude estimation rather than separate valuation of each component, this assumption is subject to challenge.

Issue B: Typical Place of Service (POS) Versus All POS

Most procedures are performed in more than one setting (hospital, physician office, ambulatory surgical centers). The times for the work components and the number of post-operative visits vary across settings. The specialty society surveys rely on vignettes to define the "typical" patient and, for surgical procedures with global periods, where the typical service is provided—e.g., in an inpatient setting. The vignettes are needed when surveys are used. Having external data available to estimate surgical times begs the question of whether a procedure should be valued based on the typical setting (i.e., where it is performed most often) or whether the values should reflect the full range of settings in which the procedure is furnished, which we term the "all POS" approach. Models 1 and 3 maintain the current framework for valuing surgical procedures by using the "typical" approach that is consistent with the current RUC process. The disadvantage is that this framework does not reflect the reality that many procedures are done in multiple settings. The all POS approach used in Model 2 potentially provides a more accurate valuation for the procedure as it reflects all sites where it is performed. It also avoids substantial changes in total RVUs in the future when the typical setting changes over time from inpatient to ambulatory setting as procedures migrate from inpatient to ambulatory settings.

Issue C: How to Value Intra-Service Intensity and Intra-Service Work

Using the BBM, a key methodological issue involves how intra-service intensity (IWPUT) should be valued, which also affects our estimates for intra-service work. As discussed in Chapter 2, intra-service work is the product of intra-service time and intensity.

$$\text{Intra-service work} = \text{IWPUT} \times \text{intra-service time}$$

We find that our RAND time estimates on average are lower than the CMS estimates. This raises the following question: How does the shorter procedure time affect the intensity of performing the procedure? We explore three options to valuing IWPUT. Model 1a assumes that a shorter procedure time translates into higher IWPUT values so that the shorter procedure time does not decrease intra-service work. The results could be viewed as an upper bound on mean IWPUT values. Model 1c assumes that intensity remains on average similar to the average intensity value derived from the CMS estimate, and therefore on average shorter procedure times translate into less intra-service work. The results could be viewed as a lower bound on mean IWPUT values. Option 1b is a blend of the other two options. It assumes that half of a shorter procedure time translates into higher intensity values and the other half translates into reduced intra-service work.

Table 3.3 summarizes how the choices on each of these issues are reflected in the models presented in this report. We have chosen these models to highlight the impact of different modeling choices. Alternative models could be created by combining the choices differently. For example, the all POS framework could be used with the different options for valuing intra-service work and intensity.

Table 3.3. Summary of Modeling Choices Reflected in Report Models

	Model 1a	Model 1b	Model 1c	Model 2	Model 3
Overview	Estimates total work for the "typical" setting using the BBM. Assumes that shorter RAND times are offset by increases in intensity so that mean intra-service work is comparable to CMS estimates.	Estimates total work for the "typical" setting using the BBM. Intra-service intensity is a blend of Model 1a and Model 1c intensity values.	Estimates total work for the "typical" setting using the BBM. Assumes that shorter RAND times do not affect intensity so that mean intra-service work is lower than the CMS estimate.	Similar to Model 1a except that the estimates are for all POS rather than "typical" setting.	Directly predicts total work for the "typical" setting. IWPUT is not separately estimated.
Issue A—Method	BBM	BBM	BBM	BBM	Single prediction model for total work
Issue B—Setting	Typical	Typical	Typical	All POS	Typical
Issue C—How to value IWPUT	$IWPUT_{RAND\text{-}P}$ = predicted intra-service work ÷ RAND time	$IWPUT_{BLEND}$ = $IWPUT_{RAND\text{-}P}$ x 0.5 + $IWPUT_{CMS\text{-}P}$ x 0.5	$IWPUT_{CMS\text{-}P}$ = predicted values based on values derived from CMS estimates	$IWPUT_{RAND\text{-}P}$ = predicted intra-service work ÷ RAND time	N/A
How are total work RVUs calculated?	Pre-service $work_{predicted}$ + intra-service $work_{predicted}$ +post-service $work_{predicted}$ + post-operative $E\&M_{predicted}$	Pre-service $work_{predicted}$ + $IWPUT_{BLEND}$ x RAND time + post-service $work_{predicted}$ + post-operative $E\&M_{predicted}$	Pre-service $work_{predicted}$ + $IWPUT_{CMS\text{-}P}$ x RAND time + post-service $work_{predicted}$ + post-operative $E\&M_{predicted}$	Pre-service $work_{predicted}$ + intra-service $work_{predicted}$ +post-service $work_{predicted}$ + post-operative $E\&M_{predicted}$	Predicted value

Overview of Prediction Models and Modeling Steps

In this section, we provide an overview of our methods for developing the prediction models for each work component. These steps are further explained in the relevant chapters of the report.

Step 1: Derive RVUs for Each Work Component

A key modeling issue is what values should be used as the dependent variables in the prediction models. One challenge that we face is the lack of an independent measure or "gold standard" for both total work and the individual work components. For components other than intra-service time, we need to rely on the CMS estimates in building our prediction models. This means that to the extent there are systematic biases in these data, we build those biases into our predicted values.

Our starting point in developing our validation models is total work RVUs. Arguably, the total work RVUs are the most accurate estimates in the current process because they receive more scrutiny than the individual components. For example, the RUC and CMS focus on the total work RVUs when they consider relativity across services and the budget neutrality calculations that apply to revised RVUs.

We use the CMS estimates for total work RVUs and the reverse BBM to derive the CMS estimates for the individual work components. Key assumptions that we make in building the model are as follows:

- The constant intensities available for pre-service, immediate post-service, and post-operative visits are accurate.
- The RUC's assumptions regarding the typical site of service for procedures with global periods are implicit in the types of post-operative visits in the CMS time file. For example, if there are inpatient E&M visits in the time file, the RUC is assuming that the procedure is typically done in the inpatient setting.
- The site of service distribution in the Medicare fee-for-service population represents the site of service distribution for the entire population. For example, if the majority of cases for a given procedure are performed in the physician office setting in the Medicare population, this is also true of those who are not covered by Medicare.

Step 2: Estimate Characteristics of Procedures Used in Prediction Models

RAND's objective for each model is to independently estimate the different components that generate a work RVU for a given service. In this step, we develop the measures for the characteristics of the procedures, including intra-service times, and measures for procedure complexity and risk and patient complexity that we use in the prediction models. Table 3.4 contains a listing of the variables that are used in the prediction models for each work component. Explanations for how we construct these variables are found in Chapter 4 for intra-service times and in Appendix C for the remaining variables.

Table 3.4. Summary of Variables Used in Prediction Models to Describe Procedure Characteristics

Variable Category	Brief Explanation	Pre-Service Model	Post-Service Model	Post-Op Model	Intra-Work Model	Intensity Model
Time for performance						
Intra-service time	How long it takes to perform the service; key variable in prediction models	X	x	x	x	x
Characteristics of procedure						
Code grouping	Codes grouped by a combination of clinical characteristics and the amount of work required	X	x	X[2]	x	x
Body system	Codes grouped by body system	X	x	X[2]	x	x
Global period	Some procedures are assigned a 10- or 90-day global period in which visits during that time for that service are paid for by the single payment	X	x	X[2]	x	x
Risk level	Categorical variable based on the least-resource-intensive setting in which Medicare covers the procedure	X	x	X[2]	x	x
Laparoscopic or thoracic procedure	May have unique aspects that impact pre-service and immediate post-service work	X	x	-	-	-
Patient and service complexity						
Comorbidities	Captured by the count of comorbidities	X[1]	x	x	x	x
Length of stay	Median length of stay in a hospital setting among those who receive the service	X[1]	x	x	x	x
ICU days	Median length of stay in an ICU among those who receive the service	X[1]	x	x	x	x
Age	Average age among Medicare beneficiaries who receive the service	X[1]	x	x	x	x
Gender	Proportion of patients female among Medicare beneficiaries who receive the service	X[1]	x	x	x	x
Major complications	Proportion of Medicare beneficiaries who receive the service who have a major complication in the subsequent 30 days.	X[1]	x	x	x	x
Mortality rate	Proportion of Medicare beneficiaries who receive the service who die in the subsequent 30 days	X[1]	x	x	x	x
Intensity						
Malpractice risk	Calculated for each specialty and individualized for a given service based on mix of specialties that bill for that service	-	-	-	x	x

Variable Category	Brief Explanation	Pre-Service Model	Post-Service Model	Post-Op Model	Intra-Work Model	Intensity Model
Years of training	Calculated for each specialty and individualized for a given service based on mix of specialties that bill for that service	-	-	-	x	x
Urgency of decisionmaking	As a marker of urgency, we measure what fraction of the services performed within the Medicare population that occur on the first day of an emergent hospital admission or in an ED followed by an admission.	-	-	-	x	x

[1] Variable was used only in pre-service evaluation regressions; it was omitted from the pre-service positioning and pre-service scrubbing regressions.
[2] Variable was not used in the regressions estimating post-operative critical care visits.

Intra-Service Times

The key variable for our models is intra-service time or "skin-to-skin" surgical time estimates derived from independent data. We have not been able to identify a single data source that covers the full range of core procedures. We use two different data sources to estimate surgical times. We use Medicare anesthesia claims to estimate times for surgical procedures that require anesthesia. Because many surgical procedures typically do not require general or regional anesthesia, we expand our coverage of ambulatory surgeries by using data from the New York Statewide Planning and Research Cooperative System (SPARCS) system. It collects patient-level detail on every ambulatory surgery performed in a hospital outpatient department or ambulatory surgical center in New York State.

As described in Chapter 4, we use both data sources to estimate intra-service time, which we term *RAND time estimates*. We calculate RAND time estimates for both the typical POS and all POS based on surgical encounters that include a single surgical procedure. We use these time estimates in the prediction models for the individual work components and total work.

Code Groupings

We recognize that groups of individual services may share characteristics that affect work values and that, therefore, grouping the codes might be useful. Because code groupings capture services with a similar approach to address similar clinical problems, it is reasonable to expect that the codes in a group would have similar values for individual work components. For example, a grouping of neurosurgical codes would likely require a similar amount of time for pre-service positioning of the patient.

The code groupings would ideally define a set of services that are clinically similar in both the problem addressed and complexity. Code groupings based just on clinical problem are less amenable to our needs. For example, placement of peripheral intravenous and replacement of indwelling catheter are both services that address a similar problem—putting an intravenous line into a vein. However, the level of complexity and resources are extremely different, as is the circumstance in which they are performed. One is performed within minutes at the bedside for most patients, and another is often done under sedation, requires monitoring, and is performed for unique circumstances, such as a need for chemotherapy. While clinically similar procedures are useful in defining groups, we would also like the codes within a code group to have similar levels of complexity and required resources.

As described in more detail in Appendix B, we explored several different options for code groupings. Our final grouping structure is a combination of two systems, CMS's Ambulatory Payment Classification (APC) and the Clinical Classifications Software (CCS) groupings developed by the Agency for Healthcare Research and Quality. We chose this hybrid system because we thought it best captured the need for groupings of procedures with a similar clinical focus and a similar level of complexity. Across the codes in our core group, there are 304 groupings.

Patient and Service Complexity

For some of our models, we include measures for the level of illness among patients who receive the service and the relative complexity of the service. These measures include patient characteristics such as age, gender, and comorbidities. For procedures performed on hospital inpatients, we examine the median length of stay and number of ICU days and the proportion of people who die within 30 days or suffer a major complication. We derive these variables from beneficiary-level Medicare claims data for inpatient and ambulatory services.

Intensity Characteristics

In the theoretical basis for the RBRVS, work RVUs are a function of the time and intensity for a given procedure. In the RUC surveys, intensity is a function of mental effort and judgment, technical skill and physical effort, and psychological stress. Some of these dimensions are captured in the variables for patient and procedure characteristics. We supplement these variables with average years of postgraduate residency training among physicians who perform the procedure as one marker of technical skill, estimated malpractice risk among physicians who perform the procedure as one marker of psychological stress, and the fraction of procedures performed on the first day of an emergent hospital admission or in the ED followed by an inpatient admission as one marker of the urgency of decisionmaking.

Step 3: Estimate Work for Pre- and Immediate Post-Service Activities (Chapter 5)

Pre-service work includes three subcomponents (evaluation, patient positioning, and scrubbing activities). Immediate post-service work includes non-"skin-to-skin" work in the operating room; patient stabilization in the recovery area; and communicating with the patient, family, and other professionals.

We do not have external time data for either pre-service or immediate post-service work. We use the CMS estimates (Table 2.1) in separate prediction models for each of the pre-service components and immediate post-service work.[19] We perform separate regressions for the typical POS and all POS. We use the predicted values from these regressions as inputs in Steps 5 and 6.

Step 4: Estimate Work for Post-Operative E&M Visits (Chapter 6)

Post-operative evaluation and management (E&M) visits related to surgical procedures are bundled into total work for CPT procedure codes with a 10- or 90-day global period. Because providers do not bill separately for post-operative E&M visits, Medicare claims data do not indicate where and how often these services are provided. We use the number and type of visits reported on the CMS physician time file to build our prediction models for codes with 10- and 90-day global periods for both the typical POS and all POS. Before doing so, we make corrections to the time file values when it appears the data are in error. For example, if the typical place of service for a procedure is the outpatient setting, but the CMS time file includes inpatient E&M visits, we remove these E&M visits. More details are provided in Chapter 6 on the corrections we apply. We use the predicted values from these regressions as inputs in Steps 5 and 6.

Step 5: Estimate Intra-Service Work and Intensity (Chapter 7)

In this step, we estimate the value of intra-service work and intensity. In Steps 3 and 4, we estimated initial work values for other work components by using the CMS time estimates for the component and standard intensity values. Intra-service work is composed of intra-service time and intensity (IWPUT). While we do estimate intra-service time in Step 2, we do not have an independent estimate of intensity. We cannot estimate intra-service work independent of the other work components and need to derive our initial estimate of intra-service work using the reverse BBM. We subtract the predicted work values for the other work components from total work RVUs.

$$\textit{Intra-service work}_{derived} = \textit{Total work} - (\textit{immediate pre-/post-service work}_{predicted} + \textit{post-operative E\&M work}_{predicted})$$

Intra-service work also equals the product of intra-service time and IWPUT. After we have predicted intra-service work, we can derive IWPUT:

$$\textit{IWPUT} = \textit{Intra-service work}_{predicted} / \textit{Intra-service time}_{RAND}$$

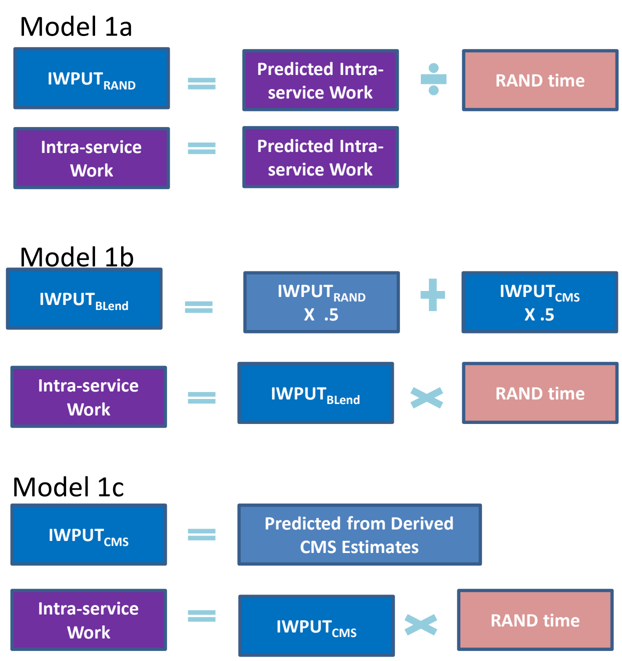
We find that the RAND intra-service time estimates are systematically shorter than CMS estimates. Using the RAND estimates without an adjustment to intra-service work increases the IWPUT values relative to values derived using the CMS estimates. The extent to which the shorter surgical times should translate into higher intensity values is not clear. As discussed in

[19] We are using the standard intensity factors associated with each pre-service and immediate post-service component.

the previous subsection, we model three options for valuing intra-service work and intensity using the BBM (shown as Issue C in Figure 3.2 and below in Figure 3.3).

- Model 1a assumes that changes in procedure times are offset by changes in intensity so that intra-service work values are unaffected by reduced intra-service times. The $IWPUT_{RAND}$ values are derived by dividing the values from the intra-service work prediction model by the RAND time estimates. The results provide an upper bound on the average intensity values.
- Model 1b is a blend of Option 1a and 1c. It assumes that half of any changes in procedure time are offset by changes in intensity and that the other half affects intra-service work. We derive the $IWPUT_{blend}$ values by averaging the Model 1a $IWPUT_{RAND}$ and Model 1c $IWPUT_{CMS}$ values. We multiply $IWPUT_{blend}$ by RAND time to estimate intra-service work.
- Model 1c assumes that reductions in intra-service time reduce intra-service work proportionately. We derive IWPUT from the CMS estimates and, after adjusting for nonsensical values, we use these estimates to predict $IWPUT_{CMS}$. We multiply predicted $IWPUT_{CMS}$ by the RAND time to estimate intra-service work. The results provide a lower bound on the mean intensity values.

Figure 3.3. Issue C Options for Valuing IWPUT and Intra-Service Work



Step 6: Estimate Total Work RVUs (Chapter 8)

Our final step is to estimate total work RVUs. At the outset, we considered whether the model should estimate total work RVUs in a single prediction model or develop an estimate by aggregating estimates of the individual work components (Issue A in Figure 3.2). We investigate both alternatives. Model 3 estimates total work RVUs in a single prediction model using the current total work RVUs and the full set of variables for patient and procedure characteristics and intensity. Under the BBM alternatives in Models 1 and 2, we aggregate the results for the individual work component predictions into a total work RVU estimate. We present the results from selected options from each of the steps as five models in Chapter 8.

Analysis of Results

At each modeling step, we analyze the results by comparing the percentage change in values between the CMS estimates and the RAND estimates. We summarize these changes by procedure characteristic and for the high-volume codes listed in Table 3.3. In the main body of the report, the summary analyses are weighted by the Medicare volume for each procedure code. In calculating Medicare volume, we use the discounted volume counts reported in the utilization file produced as part of the final rule for the 2014 Medicare physician fee schedule. The discounted volume counts account for multiple procedure discounting and bilateral procedures. We include only volumes that represent the actual performance of the surgical procedure and exclude, for example, volumes reported for assistants-at-surgery and pre-service or post-services only for procedures with global periods. Appendix H contains summary analyses for each work component without weighting by Medicare volume.

In Chapter 8, we compare the percentage change in total RVUs across the five models and include a comparison of the changes by physician specialty. In addition, we use the Spearman’s rank correlation to examine the extent to which the results change the rank order of total RVUs for the core procedures.

4. Intra-Service Times for Single Procedures

Overview of Chapter

One critical aspect of all our validation models is intra-service time. In this chapter, we focus on the intra-service times. We begin by first discussing complexities in defining intra-service time using data from external databases. Next, we introduce an analytic technique that combines existing CMS estimates and time estimates from external databases. This analytic technique helps us address situations where for a given procedure there are few observations in the external data sources. Using this technique, we use data from two data sources (SPARCS and Medicare anesthesia claims) to estimate intra-service time, what we call RAND time. Finally, we externally validate the RAND time estimates by comparing them to data from two surgery databases.

In these analyses we are focused on situations where only a single procedure was performed. Our analyses on the appropriate discounting factor for situations where multiple procedures are performed concurrently is discussed in Chapter 9 and Appendix F.

Concerns with Current Time Values

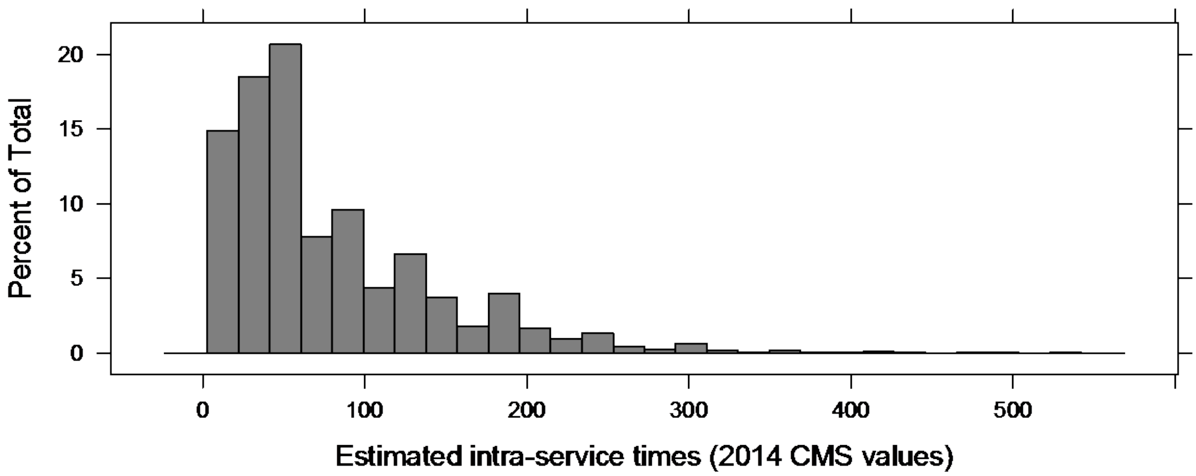
The intra-service time estimates that exist in the CMS physician time file (Figure 4.1) come from a number of sources. While some time estimates have been carried forward from the original Harvard work done decades ago, the main source of time estimates is the RUC's specialty society surveys. These survey instruments query times for procedures as described in the RUC clinical vignettes. The clinical vignettes provide background on a "typical" procedure.

Criticisms of intra-service time estimates have focused on potential biases when asking physicians to estimate how long a given procedure takes to perform. Issues include the potentially small number of survey responses (before October 2013, a minimum of 30 responses were required in order to be considered valid by the RUC)[20] and concerns that the responses may suffer from recall biases. Prior work has compared the physician estimates of intra-service time to estimates that come from external databases. For example, McCall et al. (2006) studied operating room logs and concluded that "[CMS] median intra-service time estimates are significantly longer than intra-service times from operative logs." Averaged across 60 procedures, these authors found that the estimates derived from operative logs were 31 minutes

[20] Effective October 2013, the RUC increased its requirements for each survey of commonly performed procedures based on the number of services provided to Medicare beneficiaries. The minimum requirements are at least 75 completed surveys for services provided more than 1 million times annually and at least 50 completed surveys for services provided more than 100,000 times annually (AMA, 2013).

shorter than the CMS estimates.[21] Cromwell et al. (2010) noted that the intra-service time estimates for many procedures have increased since the original estimates produced by the Harvard team. This contradicts the expectation that improvements in medical technology would lead to shorter surgical times, rather than longer times. Evaluation by Smith et al. (2007) of operative times in the Society of Thoracic Surgeons (STS) database indicates that CMS times were 7.2 percent longer than the mean operative time on a Medicare volume-weighted basis.

Figure 4.1. Distribution (Percentage) of Core Procedure Codes by Intra-Service Minutes



Data and Methods

Our objective is to estimate "typical" intra-service times; we interpret "typical" as the median time, such that half of all procedure times should be at or below the estimate and half should be at or above the estimate, or the geometric mean. For time distributions that are log-normal, the median and geometric mean of the population are the same. Both medians and geometric means provide a measure of robustness against unrealistically long time observations that may be the result of misreports or anomalous circumstances—situations some might argue should not impact payment.

As one might expect, we find substantial differences in intra-service times across place of service (POS) designations. For procedures that are performed in multiple settings, we report time estimates based on the typical POS; typical POS is defined as the most common POS for

[21] Although the RUC argued that the McCall team misstated some of the RUC and Harvard times, the RUC's correction of the McCall work still indicates that the RUC's time estimates are substantially higher than the estimates from operating room logs (Rich, 2007). The RUC also argued that the McCall results were unreliable because of differences between the measured population and the population of interest.

that procedure. We also report a weighted average of time for each setting in which the procedure is provided. As shorthand, we refer to these as the "all POS" estimates. The weights correspond to the percent of Medicare volume in each POS. Combining estimates across POS categories is a departure from the vignette approach of the RUC, where work values for procedures with a global period are elicited for a single POS, even if the procedure is routinely performed in multiple settings.

Databases

SPARCS Database

The New York Statewide Planning and Research Cooperative System (SPARCS) system collects "patient level detail on patient characteristics, diagnoses and treatments, services, and charges for every hospital discharge, ambulatory surgery patient, and emergency department admission in New York State" in order to "satisfy public health and health services administration information needs" (New York State Department of Health, 2007). The outpatient portion of SPARCS is used for both hospital outpatient departments and ambulatory surgical centers. It includes the key data element operating room (OR) time, which records the amount of time that a patient is in the operating room, exclusive of pre-operative preparation time and post-operative recovery time. The documentation indicates that "this time should have been calculated from actual entry into the ambulatory surgery operating room and should have ended at actual departure from the ambulatory surgery operating room" (New York State Department of Health, 2009). This description implies that the OR time includes positioning, which is not a portion of CMS's definition of the intra-service work, which focuses just on skin-to-skin time. Hence, we transform the OR times into estimated surgical times, as will be described below. SPARCS does not collect comparable OR times for inpatient procedures.

We note that there are several data quality issues with SPARCS that we address. First, some facilities report 1 minute of OR time for nearly all of their procedures. This suggests that reported 1-minute times may be a default response by facilities that are unwilling to report their actual times. An OR time of 1 minute is unrealistically short in any situation. We therefore remove all 1-minute observations from our SPARCS file. Secondly, some facilities in SPARCS display a high level of "clumping" of their reported times. By clumping, we mean that some facilities concentrate the majority of reported times on a small number of distinct values. Out of concerns that such data may arise from heavy rounding of reported times, we eliminate all observations from facilities where at least 50 percent of the reported time observations are distributed across five or fewer distinct time values. While some clumping may arise from regularities in OR scheduling, we believe that excluding such facilities results in a more credible data source. In all, we remove approximately 18 percent of all SPARCS observations through these exclusions.

Medicare Anesthesia Times

As SPARCS only provides time estimates for outpatient procedures, there are many surgical procedures whose times we are not able to estimate from that data source. In order to expand our coverage—as well as to validate our estimates in an alternative dataset where possible—we use Medicare anesthesia data to provide time estimates of surgical procedures that require anesthesia. Anesthesia claims are billed in complete or fractions of 15-minute increments. These data can be linked to the Medicare claims data for the surgical procedures in order to associate an anesthesia time with relevant procedure codes.

CMS gives the following definition (2014c):

> Anesthesia time is defined as the period during which an anesthesia practitioner is present with the patient. It starts when the anesthesia practitioner begins to prepare the patient for anesthesia services in the operating room or an equivalent area and ends when the anesthesia practitioner is no longer furnishing anesthesia services to the patient, that is, when the patient may be placed safely under postoperative care. Anesthesia time is a continuous time period from the start of anesthesia to the end of an anesthesia service.

Given this definition, in most cases the billed anesthesia time begins soon after the patient enters the operating room (and OR time therefore begins) and ends just before or after the patient leaves the operating room. For the same procedure, anesthesia times are actually typically somewhat longer than OR times; we adjust the times to account for these differences (Appendix D).

Different Time Measures and Patient Populations

Table 4.1 summarizes the patient populations represented in the CMS physician time file and our two primary data sources (SPARCS and billed Medicare anesthesia times) and the time elements contained in each. In the CMS time file, intra-service time estimates are reported as "median intra-service time." For most procedures, the reported times are the empirical median of the RUC survey time values. In some cases, CMS may disagree with RUC estimates. For example, in a recent response to public comments, CMS (2012b) wrote: "The AMA RUC conducted a multi-specialty survey of 110 physicians and recommended an RVU for each of the new CPT TCM [transitional care management] codes. . . . For CPT code 99496, we disagree with the observed median intra-service time of 60 minutes. We believe that 50 minutes of intra-service time is a more appropriate intra-service time for CPT code 99496." In this sense, the CMS values should be considered *estimated* rather than *empirical* medians. More generally, we consider the existing CMS estimates to be a measure of central tendency suitable for interpretation as the prevailing estimate of either the median or the geometric mean of the distribution of intra-service times.

It is important to acknowledge that the databases we use vary in what they capture in their time measures. For example, billed "anesthesia time" captures a different time component than

OR time. In turn, these times are different than what is captured by “incision” or “skin-to-skin” time. When we use multiple databases, we need to address these definitional differences.

We also acknowledge the challenge of defining “typical” when integrating data from a variety of sources and contexts. RUC surveys describe “typical” patient vignettes that may involve younger and healthier individuals than the “typical” Medicare beneficiary receiving the same procedure. We therefore explore approaches to reporting or adjusting for differences in perspective especially when combining CMS time estimates (based on RUC estimates where “typical” includes all patients) with Medicare data (which include only Medicare beneficiaries).

Because there are differences in typical intra-service times by POS, we stratify many of our analyses on this basis. We also stratify by whether the procedure was associated with anesthesia, conceptualizing use of anesthesia as another method of defining POS. Motivated by our available data sources and exploratory data analyses, our three POS designations for estimating intra-service times are Inpatient, Not Inpatient with Anesthesia, and Not Inpatient Without Anesthesia. All records that come from the Medicare data are considered to be With Anesthesia (since an anesthesiologist billed time for that case). In SPARCS, records that report general or regional anesthesia are classified as With Anesthesia; all others are classified as Without Anesthesia. Ideally, we would have an Inpatient Without Anesthesia POS category, but no external publicly available time data with sufficient coverage of these procedures are available, and it is less common to perform a procedure without anesthesia in the inpatient setting.

The various approaches to calculating central tendency estimates from individual-level databases are a related but separate point. Several statistics (including the median, arithmetic mean, geometric mean, and mode) can be used to estimate the central tendency across a distribution of individual values. The RUC survey data include the median time estimates for intra-service time. When using time data from external databases to estimate the central tendency, we must face the issue that they include some very long time observations that, based on the distribution, appear to be anomalies or errors. We explored various options for calculating central tendency, but in the end we chose to use geometric means because they confer robustness against anomalously long time observations—as medians do. Operationally, we calculate geometric means by estimating the arithmetic mean of the logarithm of the surgical time and exponentiating that estimate.[22]

While our analyses in this report are based on the SPARCS and Medicare anesthesia times, we also use data from the National Surgery Quality Improvement Program (NSQIP) and data

[22] In this regard, we note that CMS uses the geometric mean to remove statistical outliers across its prospective payment systems. Since 2013, CMS has used the geometric mean to determine relative weights under the outpatient prospective payment system. CMS changed from using median costs to establish the relative weights to the geometric mean in order to better capture the range of costs and clinical practice patterns and changes that are introduced slowly into the system (CMS, 2012a).

from cardiac surgery data from Massachusetts (Mass-DAC) to externally validate our time estimates (discussed below).

Table 4.1. Summary of Data Sources in the RAND Model

Database	Population	Procedures	Time Measurement
CMS physician time file	"Typical" patient	All surgical procedures	"Skin-to-skin" surgical time[1]
SPARCS	New York all-payer	Surgical procedures provided in a hospital outpatient, ED, or ambulatory surgical center setting with and without anesthesia	OR time[2]
Medicare Part B bills	Medicare	All surgical procedures involving anesthesia in all POS	Anesthesia time[3]

[1] Intra-service time is typically described as the "skin-to-skin" time spent in the OR.
[2] OR time is the total time in hours and minutes that the patient was actually in the operating room exclusive of pre-op (preparation) and post-op (recovery) time.
[3] Medicare anesthesia time is defined as the period during which an anesthesia practitioner is present with the patient.

Estimating Surgical Times with Anesthesia and OR Times

As highlighted in Table 4.1, CMS estimates are based on surgical time (or "skin-to-skin" time). Although neither of our external data sources captures this element directly, Silber et al. (2007, 2011) studied the feasibility of using Medicare anesthesia claims data to estimate anesthesia and surgical times that were manually abstracted from patients' charts. They found that anesthesia claim times were highly predictive of the abstracted anesthesia times (5.1 minute median absolute error) and surgical times (13.8 minute median absolute error). Silber's 2011 estimate of the formula to transform from Medicare anesthesia time to surgical time is –21.77 min + 0.805 × (anesthesia claim time in minutes). The surgical time is a fraction (80.5 percent) of the anesthesia time because anesthesia is given before the surgery begins and concludes after the procedure has finished.

Although surgical times for individual patients (e.g., as recorded in charts) may deviate substantially from the estimated values, these authors found that differences by hospital and procedure were "quite small, typically amounting to a few minutes difference by institution or procedure." The authors conclude that "the Medicare anesthesia claim can be utilized to construct an excellent measure of procedure time" and that "future investigators can feel confident that they may utilize our algorithm to better study procedure length through using the Medicare claim, without the need to collect procedure length information directly from the chart." As described in detail in Appendix D, we expand this transformation to a broader set of surgical services. We refer to this extension and update of the Silber transformation as the RAND transformation.

In analyses of data from NSQIP that has data on both anesthesia times and OR times, we find that across procedures anesthesia times are consistently slightly longer than the OR times. To address the differences in time between OR, anesthesia, and surgical times, we first estimate a transformation that transforms anesthesia times to OR times. We then transform all OR times to "skin-to-skin" surgical times.

Updating CMS Times with Data from External Sources: A Bayesian Approach

Because many CMS time estimates rely on small numbers of survey respondents, our goal is to augment the estimates with intra-service time estimates from external databases. However, for some procedures the external databases may provide only a handful of time observations. For these low-volume procedures, the resulting empirical estimates may be very poor estimates of the measure of interest (i.e., the geometric mean of the population of all intra-service times for a particular CPT code) because the small sample size results in a large amount of sampling variation. Low volume can arise for a number of reasons. For some services, the Medicare volume itself may be quite low. For other services, the volume may be high overall, but low in some places of service. In other cases, procedures may typically be performed in conjunction with other surgical services, and so there may be few single-procedure surgical encounters. We only focus on single procedures to estimate time. Lastly, our Medicare data and SPARCS data only capture procedures that are performed in conjunction with billed anesthesia or done in the OR, which might greatly decrease the number of available observations.

One mechanism to address the varying amounts of available data is to use Bayesian techniques. Under this method, when few observations are available, estimates are typically improved by "pulling" or "shrinking" the estimates toward a reasonable "prior" estimate of the true value. As a starting place for our analyses, we use the existing CMS estimates as this prior estimate. Hence, if there are only a few observations for a given procedure in the external databases, our estimates will be close to the CMS estimates. This is true even if the geometric mean of the intra-service time from the external database is substantially different from the CMS estimate. For procedures that have a large number of observations in our database, however, the prior estimate (i.e., the CMS estimate) becomes less influential, and our estimate of the population average will closely reflect the sample average values from the external databases. As an example, our external data sources provide at least 80,000 observations for each POS for CPT 43239. For this procedure, the Bayesian and raw geometric means will be practically indistinguishable. In contrast, for procedures that have an intermediate number of associated observations, the resulting estimates will be a compromise between the prior belief and the observed data. In this way, time estimates are only changed to the extent that external data provide evidence that such an update is warranted.

As described in greater depth in Appendix E, we use a Bayesian approach to estimate the intra-service time associated with each procedure. The strength of the prior information is defined to depend on the sampling variability of the true intra-service time distribution for each procedure. Hence, procedures that have little sampling variation in observed times are assumed to have less uncertainty associated with their typical time values even before observing the data than procedures whose times vary widely.

A potential drawback of this Bayesian approach occurs when the empirical estimates are systematically longer or shorter than the current CMS estimates. Indeed, consistent with the literature, the time estimates from external databases are on average shorter than the CMS estimates. If we do not address this systematic difference, the time estimates for procedures for which there is a small sample size in external databases will possibly be too long.

In order to account for this issue, we introduce an additional component to our estimates of time. Instead of simply pulling our estimates toward the CMS estimate, in what we term "adjusted Bayesian models" our "prior" estimate is the CMS estimate reduced by a constant in the log scale. This constant is estimated from the data so that procedures with few external time observations have their estimates adjusted to be in line with the estimates of the services that have many external observations. This can be important for maintaining relativity between procedures with many versus few external observations. See Appendix E for details.

Adjustments for Outliers and Different Data Sources

Our first step is to transform anesthesia times into the OR time scale. We then eliminate time observations whose log value is more than three standard deviations from the mean log time within procedure-by-POS designations as statistical outliers.[23] Outlying procedures are also eliminated in the estimation of the adjustment constants that underpin the adjusted Bayesian models. For details, see Appendix E.

After transforming anesthesia and OR times into surgical times using the RAND transformation, we find systematic differences between the times that are observed in SPARCS versus those that come from the Medicare anesthesia claims. For 15 of the 16 CCS level 1 groupings (CCS level 1 groupings classify procedures by body system), the SPARCS times are shorter, which might reflect the differences in the patient populations (SPARCS data come from all age groups, while Medicare anesthesia data come primarily from an older population). In order to be able to pool our analytic data from those two sources, we estimate an additional transformation to bring the SPARCS observations into the scale of the Medicare anesthesia observations. This transformation is defined by a constant shift in the log scale and is allowed to vary by CCS grouping. See Appendix E for details.

[23] This is consistent with the approach that CMS uses to eliminate statistical outliers under the prospective payment systems for hospital inpatient and outpatient services.

Summary of Steps to Estimate RAND Times

To summarize the discussion above, our steps to produce new time estimates are as follows:

1) Collect time data for single surgical procedures from external databases
   a) Medicare anesthesia times
   b) SPARCS OR times
2) Transform the external data times to surgical times
   a) Estimate the anesthesia to OR time transformation, as described in Appendix D using NSQIP data
   b) Estimate the RAND transformation described in Appendix D, which transforms OR time to surgical time
   c) Apply the transformation from Steps 2a and 2b to the Medicare anesthesia times and SPARCS OR times
3) Adjust SPARCS times to bring them into the scale of the Medicare anesthesia times
4) Produce Bayesian estimates for each procedure by POS grouping according to the following choices:
   a) Unadjusted prior estimates (current CMS time) versus adjusted prior estimates (current CMS time multiplied by a constant, where the constant is estimated from the difference)
   b) Geometric mean versus median
5) Produce two different POS estimates:
   a) The typical POS (the most common POS by Medicare volume for a given procedure)
   b) Combine across all POS (weighted by Medicare volume for each POS for a given procedure).

Results

In what follows, we explore the results of making different choices in steps 4 and 5 summarized above. To start, we graph the difference between the current CMS estimates and our estimated median estimates for typical POS based on how many observations we have in external datasets. This illustrates the Bayesian approach. We see from Figure 4.2 that the estimates with small available sample sizes are close to the CMS estimates in the log scale. With larger sample sizes in the empirical data for the typical POS, we see that the estimates can deviate substantially from the CMS values. When there is a difference, the Bayesian estimates tend to be shorter than the CMS values.

In Figure 4.3, we compare the geometric mean and median estimates for typical POS. In general, these estimates are close to one another, with nearly all of the points falling near the 45-degree line. Given that they are very similar, we focus on the geometric means.

In Figure 4.4, we compare the results that arise from Bayesian models that pull the estimates toward the unadjusted versus adjusted CMS time estimates. Since, across services, the external time estimates tend to be shorter than the CMS estimates in every POS, estimates based on adjusted CMS times are, without exception, no larger than those that are pulled toward the unadjusted CMS estimates. The values in this figure for the zero sample size indicate the typical differences between our empirical estimates and the CMS values: On average, the Inpatient estimates are 94.2 percent as long as the CMS estimates, 79.8 percent as long for Not Inpatient with Anesthesia, and 74.2 percent as long for Not Inpatient Without Anesthesia. The top "curve" in this figure corresponds to the estimates for the Inpatient POS, and the bottom corresponds to procedures whose typical POS is Not Inpatient Without Anesthesia.

In Figure 4.5, we display a comparison of our RAND time estimates to the CMS values for all POS. Although for the majority of procedures RAND times are shorter than the CMS values (the points under the 45-degree line), around 16.6 percent of the estimates are longer.

Figure 4.2. Difference in Log Scale between CMS and Bayesian Estimates (Unadjusted Prior) by Number of Observations in SPARCS Data for Typical POS

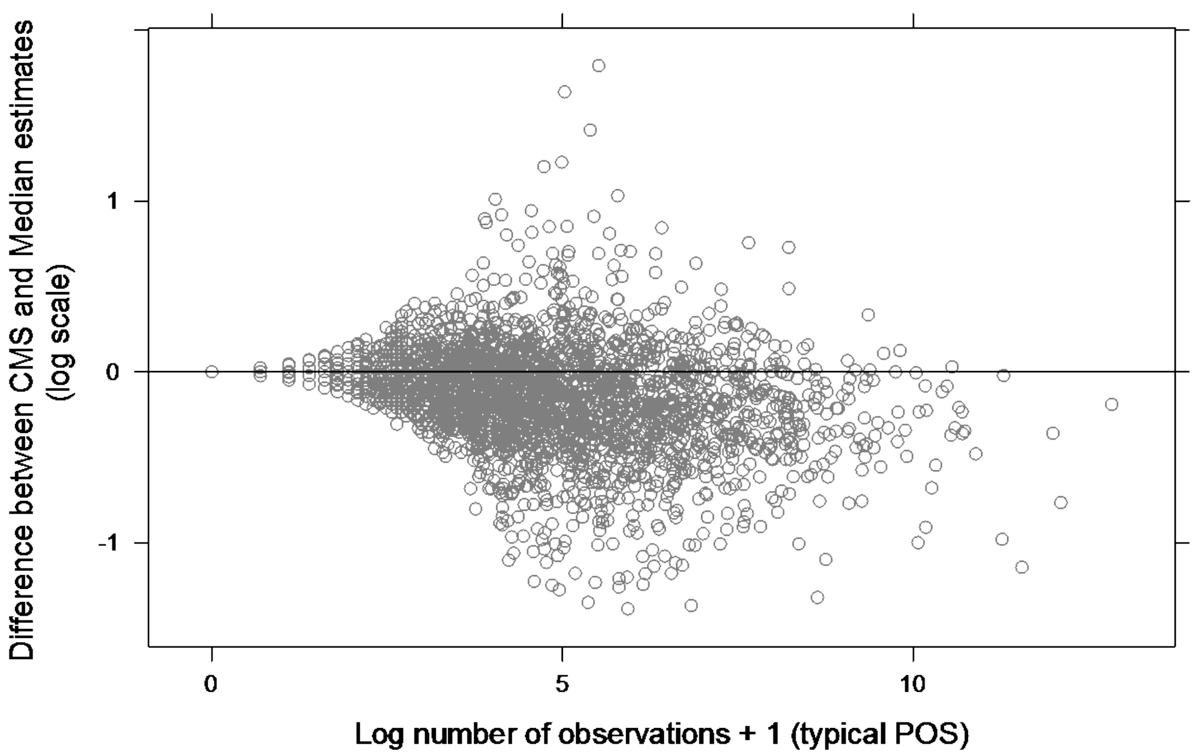


Figure 4.3. Bayesian Median Estimates Versus Geometric Mean Estimates (Typical POS and Unadjusted Prior)

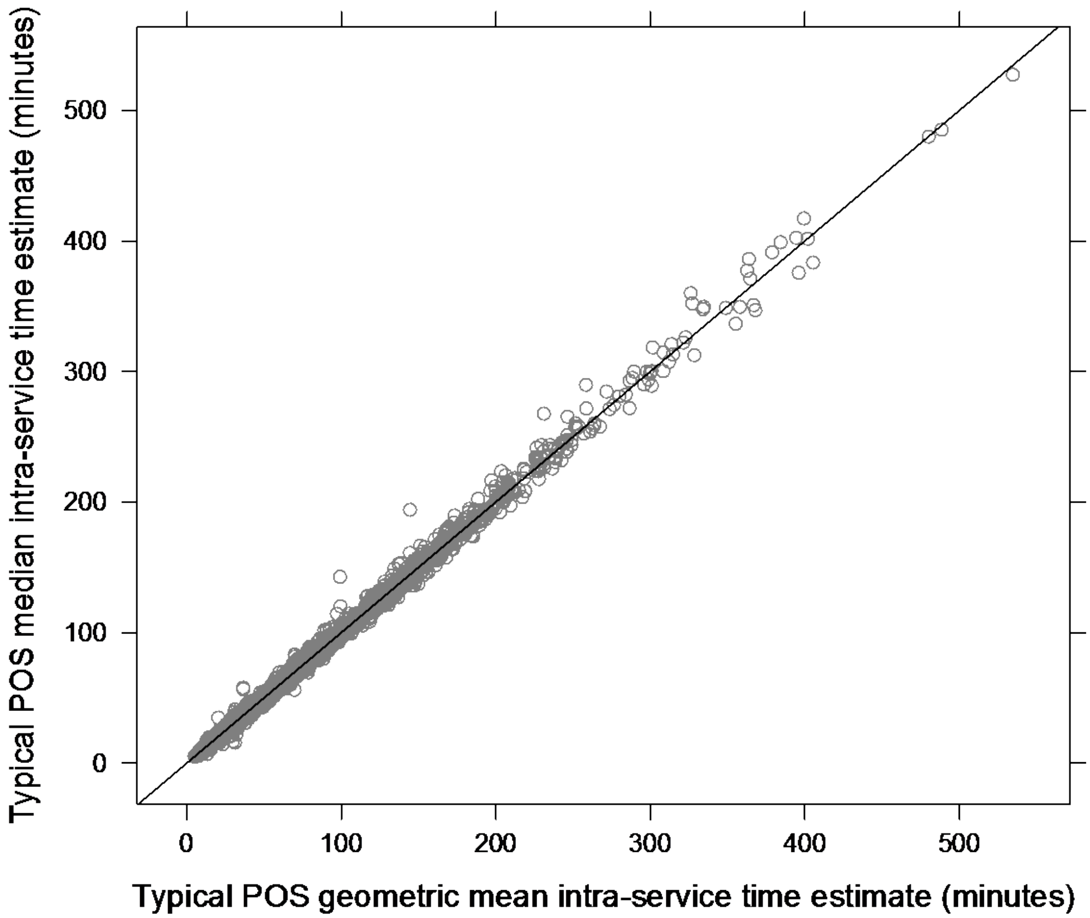
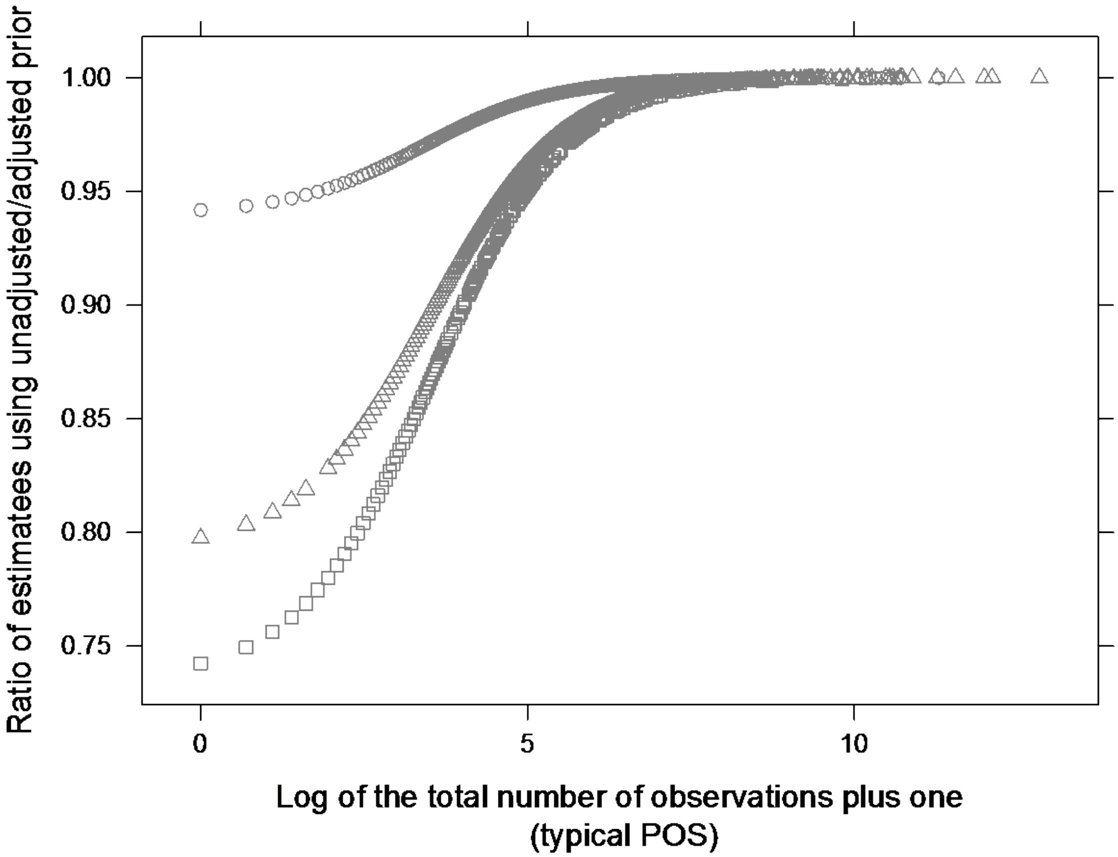
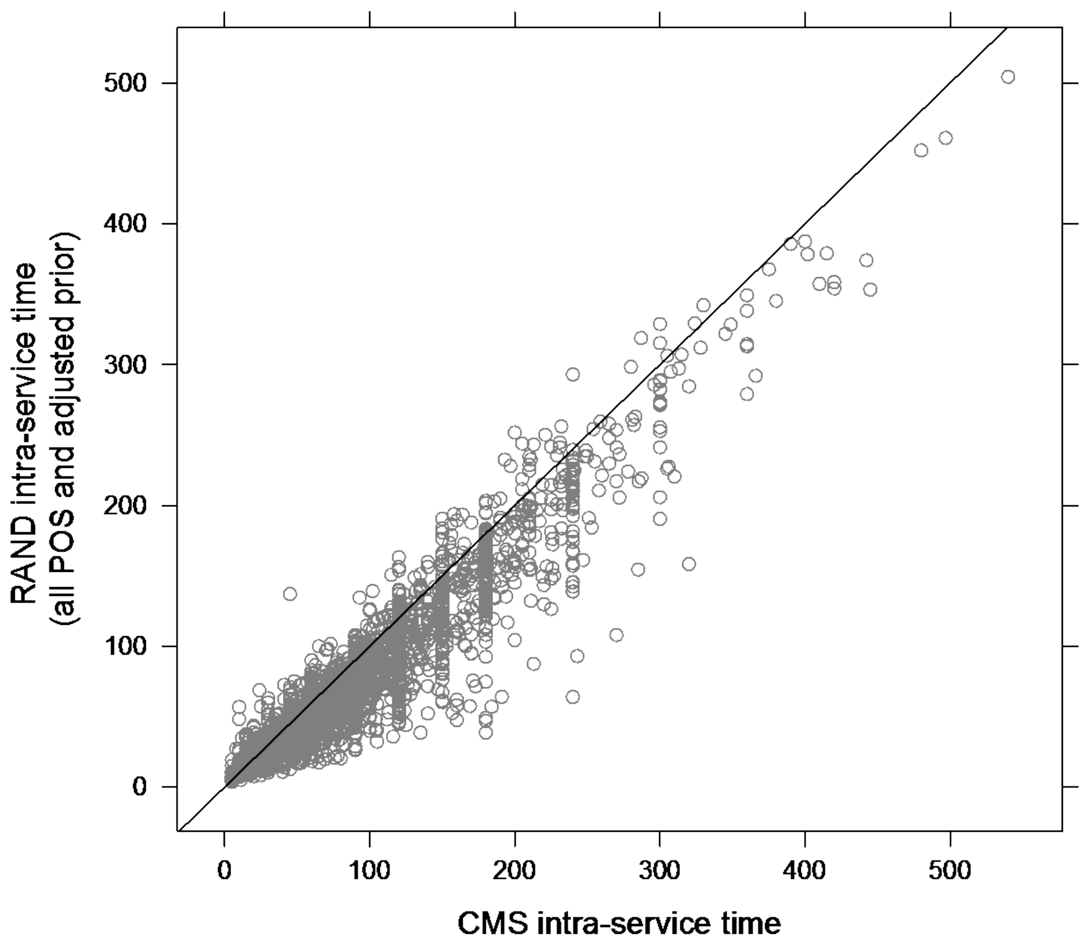


Figure 4.4. Ratio of Estimates Arising from the Adjusted to the Unadjusted Prior Using the Geometric Mean and Typical POS



NOTES: Circles represent Typical POS of Inpatient, Triangles represent Not Inpatient with Anesthesia, and Squares represent Not Inpatient Without Anesthesia.

Figure 4.5. RAND Time Estimates (All POS and Adjusted Prior) Compared to CMS Values



We present comparisons for our RAND times stratified by several key classifications in Table 4.2. All of these numbers are weighted by 2011 Medicare volume. For shorter procedures (as measured by the CMS time estimates), our updated estimates are somewhat longer on average. For longer procedures (at least 30 minutes), the opposite is true. The greatest reduction (in a relative sense) is for the procedures of intermediate length, between 31 and 70 minutes. This is echoed to a certain extent in our comparison of office versus inpatient procedures.

On average, our RAND time estimates of procedures typically done in an office are longer in the unadjusted Bayesian models, while our RAND time estimates of procedures typically done in an inpatient setting are shorter. The qualitative difference between the unadjusted and adjusted Bayesian models for this POS designation seems to reflect relatively weak coverage of some office-based procedures in our external data sources. The results for the ED POS category are somewhat anomalous but reflect the small number of codes in this category and the weighting by volume. If the results were not weighted by Medicare volume, the times for ED procedures would be essentially unchanged (results not shown).

Table 4.3 gives estimates for our "Top 20" procedures. In general, the various RAND estimates are relatively close to one another but can deviate substantially from the CMS estimates. For example, the CMS estimate for CPT 45380 is 51.5 minutes, whereas the RAND estimates are under 20 minutes.

Table 4.2. Percentage Difference Between CMS Time and RAND Time Estimates, by Procedure Category

	CMS Time Estimate Weighted Mean	Median, Prior Based on Unadjusted CMS Times, Typical POS	Geometric Mean, Prior Based on Unadjusted CMS Times, Typical POS	Geometric Mean, Prior Based on Adjusted CMS Times, Typical POS	Geometric Mean, Prior Based on Adjusted CMS Times, All POS
	Minutes	% Difference	% Difference	% Difference	% Difference
Total	24.6	–9.5	–8.7	–14.5	–14.2
CMS intra-service time categories					
0 to 30 minutes	11.0	18.4	20.5	10.2	11.2
31 to 70 minutes	42.6	–27.8	–27.4	–30.8	–30.9
71 to 120 minutes	97.9	–19.4	–19.0	–24.7	–24.2
Over 120 minutes	174.5	–17.5	–18.0	–19.2	–19.3
Global period					
0	16.5	–1.0	–0.3	–9.2	–9.4
10	25.1	–12.4	–10.2	–18.1	–16.3
90	67.3	–19.1	–18.5	–20.1	–19.2
Not applicable	18.9	–21.9	–19.8	–25.5	–26.1
Typical place of service					
ED	19.0	80.8	86.1	69.3	70.1
Inpatient	58.8	–9.0	–8.1	–9.3	–12.0
Office	13.9	8.7	10.1	–4.1	–2.5
Ambulatory facility (HOPD or ASC)	33.0	–29.5	–29.4	–31.1	–29.4
Risk category					
Office-based	12.1	23.5	25.6	9.7	10.3
ASC	33.3	–26.9	–26.6	–29.0	–27.8
Hospital outpatient	52.7	–19.2	–19.3	–21.0	–25.6
Inpatient only	132.3	–7.3	–7.0	–7.8	–8.1
Body system grouping					
Nervous system	19.7	–1.8	–2.0	–6.1	–7.9

	CMS Time Estimate Weighted Mean	Median, Prior Based on Unadjusted CMS Times, Typical POS	Geometric Mean, Prior Based on Unadjusted CMS Times, Typical POS	Geometric Mean, Prior Based on Adjusted CMS Times, Typical POS	Geometric Mean, Prior Based on Adjusted CMS Times, All POS
	Minutes	% Difference	% Difference	% Difference	% Difference
Endocrine system	101.5	–24.9	–24.4	–25.1	–22.4
Eye	17.4	13.7	15.2	10.8	12.9
Ear	11.2	22.9	35.4	18.5	18.6
Nose, mouth, and pharynx	20.0	–5.6	–2.4	–14.5	–12.3
Respiratory system	23.1	–14.5	–8.3	–12.7	–13.2
Cardiovascular system	48.0	–8.0	–8.7	–11.4	–13.7
Hemic and lymphatic system	67.1	–11.7	–11.9	–13.2	–14.0
Digestive system	39.3	–41.1	–40.4	–41.0	–40.8
Urinary system	26.4	–19.5	–14.9	–18.0	–14.6
Male genital organs	45.7	–17.1	–16.7	–18.6	–17.8
Female genital organs	40.1	–9.4	–8.7	–14.0	–14.6
Musculoskeletal system	22.5	5.4	6.2	3.6	3.9
Integumentary system	19.6	–6.3	–6.1	–20.4	–18.4
Miscellaneous services	13.1	6.7	6.0	–9.5	–8.8
2012 Medicare annual volume					
Less than 1,000	92.5	–10.9	–11.0	–15.2	–15.5
1,000 to 9,999	67.0	–16.5	–16.2	–19.4	–19.2
10,000 to 99,999	40.9	–18.2	–17.7	–22.4	–21.4
100,000 or more	15.5	–0.1	1.2	–6.7	–6.7

Figure 4.6 Comparison of All POS RAND Intra-Service Time Estimates and NSQIP Median Surgical Times, as a Function of the Log Number of NSQIP Observations Available

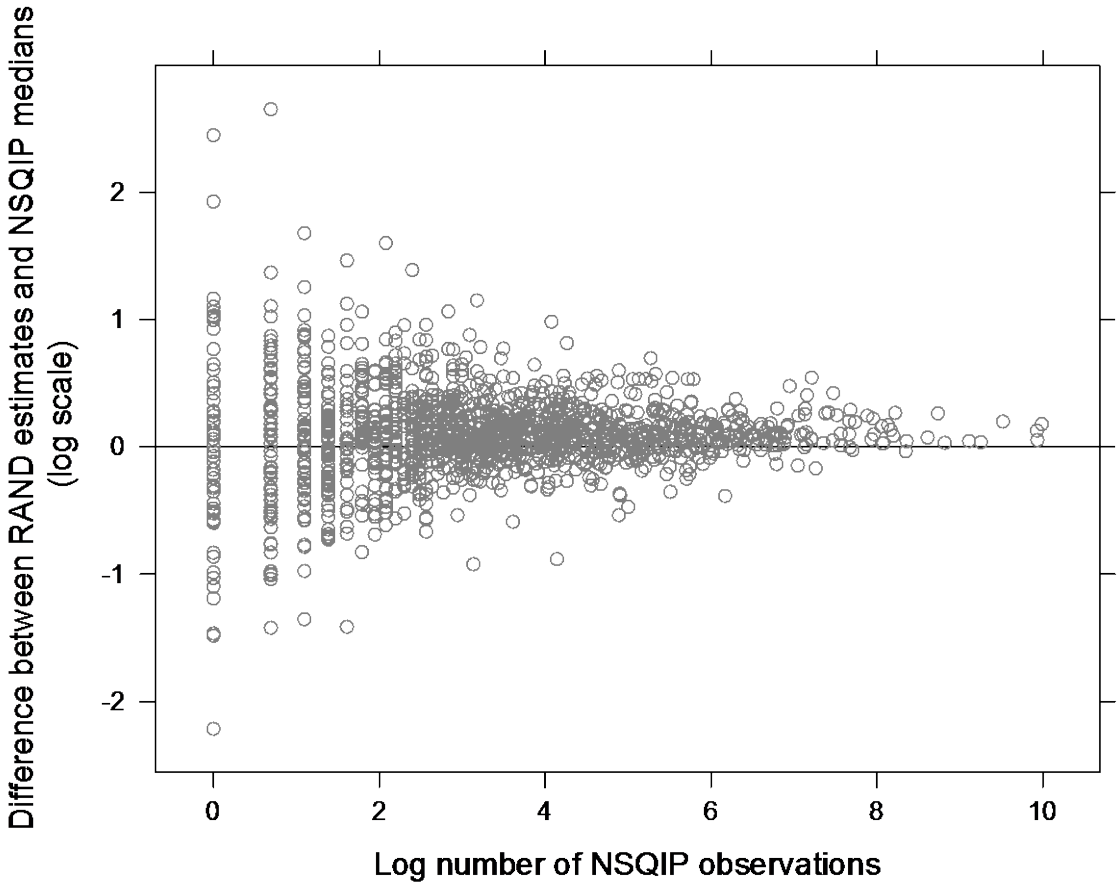


Figure 4.7. Comparison of RAND Estimates and NSQIP Mean Surgical Time Estimates, as a Function of the Log Number of NSQIP Observations Available

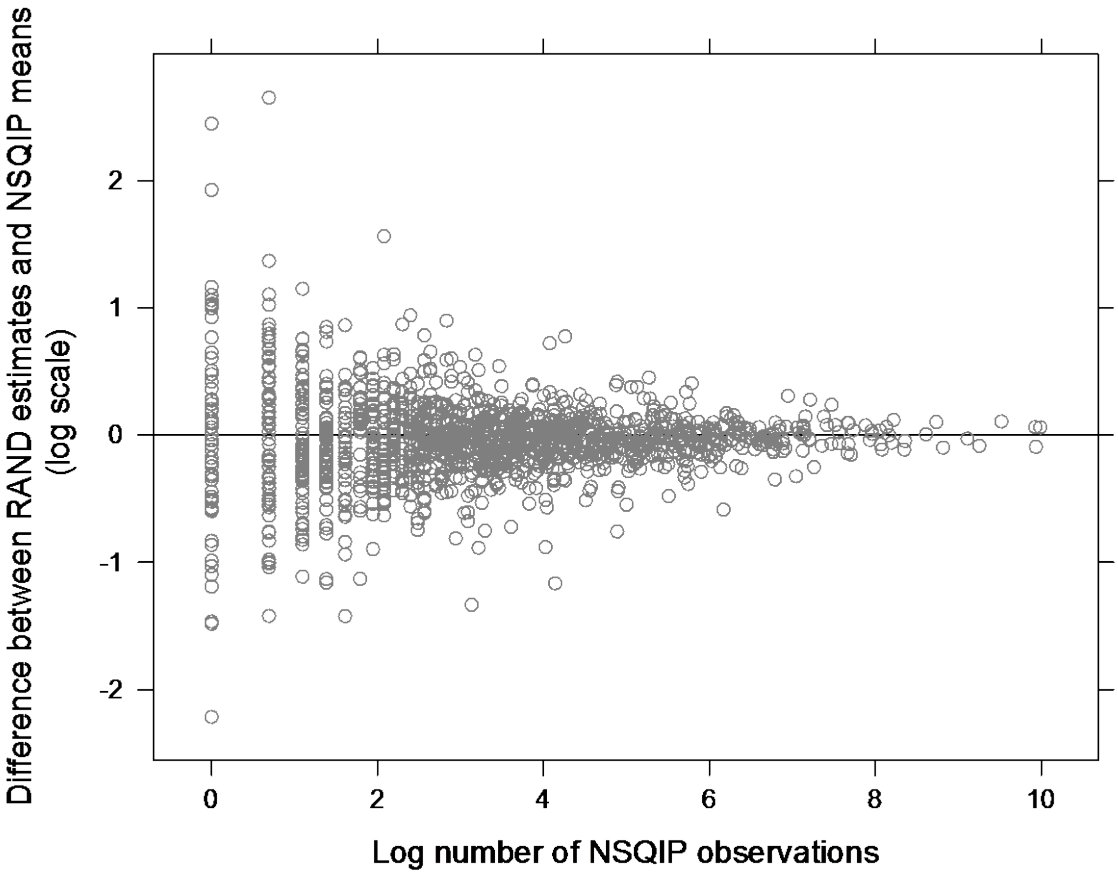


Table 4.3. CMS and RAND Intra-Service Time Estimates for "Top 20" Procedures

Code	Description	CMS Time Estimate Weighted Mean	Typical Setting Median, Prior Based on Unadjusted CMS Times	Typical Setting Geometric Mean, Prior based on Unadjusted CMS Times	Typical Setting Geometric Mean, Prior Based on Adjusted CMS Times	All POS Geometric Mean, Prior Based on Adjusted CMS Times
13132	Complex laceration repair	50	40.4	38.7	32.9	33.3
17311	Mohs procedure of skin	110	110	110	81.7	81.3
20610	Drain or inject a bursa or joint	5	10.6	10.6	10.5	10.6
27245	Treat thigh fracture	80	65.0	68.1	68.1	68.1
27447	Total knee arthroplasty	100	97.9	98.8	98.8	98.9
31231	Diagnostic nasal endoscopy	7	7.5	8.2	6.5	6.6
33533	CABG arterial single	158	192.2	194.1	193.7	193.3
35301	Rechanneling of artery	120	110.5	111.0	111.0	111.0
43239	Esophagogastroduodenoscopy (EGD) with biopsy	15	10.5	10.7	10.7	10.1
44120	Removal of small intestine	134	92.8	95.7	95.7	95.6
45380	Colonoscopy with biopsy	51.5	16.4	16.9	16.9	16.7
47562	Laparoscopic cholecystectomy	80	49.5	51.3	51.3	57.0
52000	Cystoscopy	15	10.3	10.3	10.3	11.1
52601	Prostatectomy (TURP)	75	49.9	51.8	51.1	53.7
62311	Inject spine lumbar/sacral	10	11.3	10.3	10.3	10.0
63047	Remove spinal lamina (lumbar)	90	94.1	94.2	94.1	92.3
64450	Digital nerve block	5	11.3	11.5	10.7	9.5
66984	Cataract surgery with IOL lens (1 stage)	21	17.4	17.4	17.4	18.2
67228	Treatment of retinal lesion	60	34.8	29.9	29.7	29.5
93458	Left heart artery and ventricle angiography	45	33.4	35.2	35.1	27.4

External Validation

Where possible, we have made an effort to validate the RAND time estimates to surgical time estimates from other external datasets. The advantage of these other datasets is that they directly collect surgical time. The disadvantage of these other datasets is that they are not publicly available (or are less accessible), do not include data on all surgical services, and, with the exception of NSQIP, do not use CPT codes. We therefore chose not to use these databases in our estimates of RAND time and rather used them as means of external validation.

One validation data source is the NSQIP data. For this comparison, we look at differences between the RAND (geometric mean with all POS with adjusted CMS times) estimates and the median values for the 1,399 of our core codes that appear as standalone procedures in the NSQIP data. We then plot these differences in the log scale. The NSQIP estimates are raw medians, so—when the sample size is small—the values may not be close to the population median. We see in Figure 4.6 that when the NSQIP sample size is large the RAND time estimate and the NSQIP time are similar. On average, the NSQIP values are slightly shorter than the RAND estimates (i.e., more points are above the horizontal line than below).

Interestingly—though perhaps coincidentally—the *mean* NSQIP values are even closer to the RAND values than the median NSQIP values are, even though the geometric mean underlying the RAND estimates should behave more like a median (Figure 4.7). All in all, we take this analysis of external data (i.e., data that were not directly used in the time calculations) to support the idea that the methodology—especially the RAND transformation—used to produce the time estimates appears reasonable.

We also compare the RAND time estimates for cardiac surgery to time values from Mass-DAC.[24] All Massachusetts hospitals that perform cardiac surgery are required to submit information, including time elements, to Mass-DAC. Due to data privacy considerations and logistical issues, we were only able to access summary data for procedures that had sufficient volume in the Mass-DAC data. Although this data source only gives information for a small number of CPT codes, some have relatively high volumes in the Medicare program, and therefore they represent a significant slice of Medicare spending.

[24] We also approached the Society of Thoracic Surgeons about using their database for cardiothoracic surgery for this project. While the society was willing to consider our request for validating the RAND times, there might have been restrictions on what could publicly be presented. Given time and resource constraints and concerns about potential transparency, we did not obtain the data for this project.

Table 4.4. Comparison of Cardiac Surgery Intra-Service Times from Mass-DAC and RAND Times; Sample Sizes (N) Are Given in Parentheses

CPT	Mass-DAC Median (N)	Mass-DAC Mean	RAND (N)	CMS
33533	158 (113)	165	191.3 (811)	158
33534	247 (33)	262	232.2 (164)	193
33430	271 (145)	277	255.8 (551)	232
33405	208 (885)	218	227.8 (3567)	197

The results are given in Table 4.4. In contrast to most other surgical services, it is unusual that for these services RAND times are longer than the CMS times. However, we see that in most cases the median Mass-DAC times are even longer. Some of this difference between RAND and Mass-DAC times is attributable to the Bayesian nature of the RAND estimates, since the CMS times are generally shorter than the empirical observations for these CPT codes. Also important is that the Mass-DAC times come only from Massachusetts. The RAND times are typically between the CMS and Mass-DAC estimates. The two services for which the RAND estimates are not between the CMS and Mass-DAC values—CPT 33533 and 33405—are not substantially longer than the other estimates.

Conclusion and Implications for Our Model

In summary, our data-adjusted Bayesian estimates are typically shorter than the current CMS estimates. These results are compatible with previous research that has found that CMS time estimates tend to be somewhat longer than observed times found in empirical datasets (McCall, Cromwell, and Braun, 2006; Rich, 2007, Smith et al., 2007; Cromwell et al., 2010).

In many scientific applications, giving substantial weight to existing estimates—as we do here with the CMS estimates—can be controversial. This controversy arises in part from the fact that there may be little consensus on *which* existing information should be incorporated into analyses of new data. In our case, even if there is disagreement as to whether the CMS times are accurate, they still serve as a useful default, and changes to these estimates arguably require empirical evidence. The Bayesian approach works in just this way: If little empirical information is available, the benchmark estimates will hold. (In the case of the adjusted Bayesian models, the benchmark estimates hold in a relative rather than absolute sense.) As more information accrues, however, the prior information (i.e., the current or adjusted CMS intra-service time estimates) becomes less influential. And, if new data become available over time, the estimate at one period can become the prior information for the next period, allowing for a continual learning process whereby the new data sharpen old estimates.

Incorporating Results into the Model

In this chapter, we examine alternative ways to derive RAND time estimates using the Bayesian approach and summarize the results in Table 4.2. In remaining chapters of this report, we use the RAND time estimates presented in the last two columns of Table 4.2 to estimate the other work components. The first estimate reflects our best estimate for surgical times in the typical POS. The second reflects our best estimate for all POS. Both estimates use the geometric mean and adjusted CMS times. Although the typical POS time estimates may be easiest to implement for the vignette style of eliciting times from physicians, our more data-driven approach to estimating these quantities makes the POS-weighted average feasible. Since there are systematic differences in time estimates by POS, we believe that accounting for the actual distribution of procedures across settings may produce more realistic estimates of physician work.

Limitations

As outlined above, we face several difficulties that make our estimates less reliable than those that could be derived if optimal data were available. If we had a single data source that reflected the patient population of interest and directly recorded surgical time elements by CPT code, we would be highly confident in the fidelity of our updated estimates. The available data deviated from this ideal in several ways. First, the primary data sources recorded billed anesthesia times and OR times. To address this difficulty, we extended the work of Silber et al. (2011) in order to transform OR or anesthesia times into "skin-to-skin" surgical times.

A second difficulty is that no available data source fully captures the demographics of the Medicare patient population and where this population receives a given procedure. This difficulty was overcome in two main ways. First, we stratified analyses by three POS designations (Inpatient, Not Inpatient with Anesthesia, and Not Inpatient Without Anesthesia). By using the typical Medicare POS or weighting estimates by POS distributions, our estimates reflect prevailing practice patterns rather than the potentially idiosyncratic POS mixes in our data sources. Relatedly, the use of Bayesian techniques leverages existing CMS estimates to provide reasonable estimates to combinations of POS and procedure that are poorly covered in our external data sources. Second, we adjusted for differences between patient populations in our two main data sources by bringing the SPARCS observations into a scale defined by the Medicare anesthesia observations. In almost all cases these adjustments lengthened the SPARCS observations, meaning that this adjustment is conservative in the sense of erring on the side of longer time estimates. However, we are unable to assess how a procedure that is performed in an office setting compares to those that are performed in facility settings or with anesthesia.

Even with these difficulties and shortcomings, we believe that these time estimates—in aggregate—may be more reliable than the current CMS time values. CMS time estimates have been found to be too high in numerous empirical studies; our work confirms these findings. And, over time as better data sources become available, we believe that the analytic methods outlined here can be used to substantially improve on previous procedure-by-procedure estimates of intra-service times. Our methods provide a methodology and path to move away from time data that are based entirely on surveys by building on existing CMS time estimates. Our adjusted Bayesian model is able to update intra-service times where we have sufficient external data while still maintaining relativity between high- and low-volume procedures (where volume is measured in the currently available data sources).

5. Pre-Service and Immediate Post-Service Work

Overview

Under the RBRVS, physicians' activities immediately before or after intra-service ("skin-to-skin") effort are bundled with payment for the procedure itself. In the case of surgical procedures, the four pre-service and immediate post-service activities include (1) patient evaluation, (2) positioning, and (3) scrubbing, as well as the (4) immediate post-operative care. This chapter describes these four pre-service and immediate post-service work components and develops models that can be used to identify procedures with more or less pre-service or immediate post-service work than is expected based on observable characteristics of the procedure, including RAND intra-service time. The time values for the four activities in the CMS estimates are multiplied by a set of constant intensity values to calculate pre-service and immediate post-service work. We explore the relationships between the CMS estimates associated with pre-service and immediate post-service components, and we describe how pre-service and immediate post-service work can be predicted. Post-operative E&M visits are a separate component of work under the BBM and are addressed separately in Chapter 6.

Pre-Service and Immediate Post-Service Subcomponents

Pre-service work and immediate post-service work are divided into four categories in the CMS time file: (1) pre-service evaluation work, (2) pre-service positioning work, (3) pre-service scrubbing and other miscellaneous activities involving physician work (which we refer to as "pre-service scrub"), and (4) immediate post-service work. Each procedure has up to four individual time values listed in the CMS time file that correspond to these four categories.

Each of the four categories has an intensity value that is constant across all procedures that can be used to convert the time values to work RVUs. In other words, a minute of work in one of the pre-service or post-service components (for example, the pre-service evaluation component) has the same intensity, regardless of such factors as where the procedure was performed. The constant intensity values were developed in the original RBRVS studies. Table 5.1 lists the pre-service and immediate post-service work categories and intensities. It also shows the proportion of core procedures with nonzero time reported for the category and reports both the mean time among codes with nonzero time and the work associated with the mean time, calculated by multiplying the mean time by the constant intensity.

Table 5.1. Pre-Service and Immediate Post-Service Work Components, Intensity, and Descriptive Statistics

Work Component	Constant Intensity	Percentage of CPT Codes in Our Core List with Nonzero Time for This Category	Mean Minutes for Codes with Nonzero Time	Mean RVUs in This Category Among Codes with Nonzero Time (Product of Mean Minutes and Constant Intensity)
Pre-service evaluation	0.0224	98.8%	33.7	0.75
Pre-service positioning	0.0224	36.8%	10.0	0.22
Pre-service scrub	0.0081	61.1%	16.2	0.13
Immediate post-service	0.0224	99.1%	25.6	0.57

Correlation Between Categories

The times associated with pre- and immediate post-service categories are, for the most part, positively correlated with one another and with the RAND intra-service time estimates (Table 5.2). Pre-service evaluation time is positively and highly correlated with intra-service time and immediate post-service time (coefficients of 0.69 and 0.67, respectively). In contrast, pre-service scrub and pre-service positioning times are less correlated with intra-service time and the other categories. In the case of pre-service scrub time, the correlation with pre-service evaluation time and immediate post-service time is negative.

Table 5.2. Correlation in CMS Times Across Pre and Post-Service Categories and RAND Intra-Service Time Estimates for Average Setting

Category	RAND Intra-Service Time	Pre-Service Evaluation Time	Pre-Service Positioning Time	Pre-Service Scrub, Etc., Time	Immediate Post-Service Time
Adjusted intra-service time	1.00				
Pre-service evaluation time	0.69	1.00			
Pre-service positioning time	0.27	0.07	1.00		
Pre-service scrub, etc. time	0.16	–0.15	0.27	1.00	
Immediate post-service time	0.55	0.67	0.05	–0.05	1.00

Prediction Models for Pre-Service and Immediate Post-Service Times

We developed a series of models predicting pre-service and immediate post-service work as a function of procedure-level observables. Specifically, we fit separate regression models that estimate the length (in minutes) of each of the four pre-service and immediate post-service categories as a function of procedure-level observables. We focus on time rather than work for two reasons. First, the underlying CMS estimate is in terms of minutes rather than RVUs. Second, because time in minutes is converted to work in RVUs using a constant intensity factor, we would obtain the same regression results regardless of whether we used time or work as a dependent variable.

We generate predicted pre-service and immediate post-service times using the estimated coefficients from these models. The difference between the predicted values and the values currently in the CMS time file (i.e., the residual) is a helpful tool to identify procedure codes that are associated with more or less pre-service and immediate post-service work than expected based on our predictive models.

We estimate two sets of models, one set for the typical POS and the other set for all POS. In each set there are four models, one for each of the following: (1) pre-service evaluation time, (2) pre-service positioning time, (3) pre-service scrub time, and (4) immediate post-service time.

For three components (pre-service evaluation, pre-service positioning, and immediate post-service time), we fit models of log minutes in each category as a function of the logged updated intra-service time estimate, CPT-level explanatory variables calculated from CMS data, code grouping random effects, and an error term. If the code groupings entered the model as fixed effects, it would be redundant to include the CCS level 1 fixed effects. In the random effects model, however, this is a consequential choice. In particular, random effects that are estimated with little external information will yield estimates that are pulled toward the CCS level 1 estimate rather than the global mean, as would be the case if the level 1 fixed effects were excluded. Because we expect substantial heterogeneity between the level 1 grouping factors, we include them in the model.

We fit these models as a gamma generalized linear mixed model with a log link function (with Stata 13's "meglm" procedure), which accommodates both fixed and random effects. The specific explanatory variables are listed in Table 5.3. See Appendix C for a detailed description of the specific independent variables. Summary statistics for the dependent and independent variables are reported in Table 5.4 (with RAND time and patient characteristics calculated for the typical POS).

The distributions of time file pre-service evaluation, pre-service positioning, and immediate post-service times each have a long right tail with a small number of extreme values. For example, 12 CPT codes (<0.5 percent of the core set of procedures) were associated with more than 140 minutes of pre-service evaluation time, with a maximum of 225 minutes for CPT 61605 (resect/excise cranial lesion). While the log link mitigates the impact of these outliers in our main results, an alternative approach is truncation. We explored this method by truncating unlogged continuous dependent variables at three standard deviations above the mean time value to reduce the influence of outliers on estimated coefficients. Truncation affected 1.4, 0.7, and 2.0 percent of pre-service evaluation, pre-service positioning, and immediate post-service procedure codes, respectively. Results from the truncated models were not substantively different from the main models.

For the fourth component (pre-service scrub times), we used a different estimation approach. In the CMS estimates, pre-service scrub times clustered into multiples of five minutes: 0, 5, 10, 15, 20, and 25 minutes. Only 2 percent of CPT codes in our core list have a pre-service scrub time value other than one of these values. In our prediction model, we first round the CMS reported times up to the nearest five-minute increment, capped at a maximum of 25 minutes. We then define five groups for 5, 10, 15, 20, and 25 minutes of pre-service scrub time, respectively, and fit a mixed effects ordered logit model predicting the ordinal dependent time variable as a function of the explanatory variables listed in Table 5.3.

Table 5.3. Variables Included in Pre-Service and Immediate Post-Service Models

Variable	Description
Estimation sample	Procedure codes with pre-service or immediate post-service times in the relevant category
Dependent variables	Time associated with the pre-service or immediate post-service component from the CMS time file
Independent variables*	
RAND time estimates (log)	Separate intra-service time estimates for typical POS and all POS
Procedure characteristics	Other characteristics of the procedure that may affect pre-service and immediate post-service work.
Patient complexity	Measures of patient characteristics and complexity calculated separately for typical POS and all POS that are used in the pre-service evaluation and immediate post-service prediction models. These variables are not included in the pre-service positioning and pre-service scrub models.

* Details on independent variables are described in Chapter 3 and Appendix C.

We fit the models using only procedures with nonzero times reported for the outcome of interest. For all procedures with CMS-reported times, we predict a series of new time values using the estimated model coefficients, including random effects (eight predicted values in total, one for each pre-service and post-service component under the typical POS and all POS approaches). The model results discussed below are from the typical POS set of regressions. We discuss the all POS regressions at the end of this chapter. We incorporate the predicted values from these regressions in our integrated models (see Chapter 7).

Table 5.4. Descriptive Statistics, Pre-Service and Immediate Post-Service Models, Typical Place of Service

A. Values for Continuous Variables

	Pre-Service Evaluation N=3,141				Pre-Service Positioning N=1,171				Pre-Service Scrub N=1,942				Immediate Post-Service N=3,150			
Variable	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
Dependent variable (mins)	33.7	24.1	1.0	225.3	10.0	6.9	1.0	45.0	16.2	7.2	1.0	30.0	25.6	27.5	2.0	401.0
RAND intra-service time	67.1	60.0	2.2	504.4	72.5	62.3	4.6	387.5	66.9	54.6	4.6	387.5	67.0	60.0	2.2	504.4
Median length of stay	2.0	2.9	0.0	18.0	2.1	3.0	0.0	18.0	2.0	2.9	0.0	18.0	2.0	2.9	0.0	18.0
Median ICU days	0.5	1.5	0.0	19.0	0.7	1.6	0.0	15.0	0.5	1.5	0.0	19.0	0.5	1.5	0.0	19.0
% of patients female	0.5	0.2	0.0	1.0	0.5	0.2	0.0	1.0	0.5	0.2	0.0	1.0	0.5	0.2	0.0	1.0
Median patient age	70.5	5.2	30.0	85.0	70.9	5.0	35.0	84.0	70.6	5.0	32.0	85.0	70.5	5.2	30.0	85.0
Median no. of comorbidities	1.5	0.9	0.0	5.0	1.7	0.9	0.0	5.0	1.6	0.9	0.0	5.0	1.5	0.9	0.0	5.0
Complication rate	0.1	0.1	0.0	0.8	0.1	0.1	0.0	0.7	0.1	0.1	0.0	0.8	0.1	0.1	0.0	0.8
Mortality rate	0.0	0.1	0.0	0.9	0.0	0.1	0.0	0.9	0.0	0.1	0.0	0.7	0.0	0.1	0.0	0.9
Laparoscopic flag	0.0	0.2	0.0	1.0	0.0	0.2	0.0	1.0	0.0	0.1	0.0	1.0	0.0	0.2	0.0	1.0
Thoracic flag	0.0	0.2	0.0	1.0	0.1	0.3	0.0	1.0	0.0	0.2	0.0	1.0	0.0	0.2	0.0	1.0

B. Distribution of Codes by Categorical Variables

Variable Category	Preservice Evaluation Percentage of Procedures (%)	Pre-Service Positioning Percentage of Procedures (%)	Pre-Service Scrub Percentage of Procedures(%)	Immediate Post-Service Percentage of Procedures(%)
Global period				
0 days	20.6	28.7	20.2	20.7
10 days	11.6	9.4	7.2	11.6
90 days	67.3	61.8	72.3	67.2
XXX	0.5	0.5	0.3	0.6
Typical place of service				
Ambulatory surgical Center	9.9	7.4	12.6	9.9
Emergency department	2.1	1.9	1.8	2.1
Inpatient hospital	40.0	40.8	41.0	40.2
Office	16.6	12.2	8.8	16.7
Outpatient hospital	31.0	37.6	35.6	31.0
Other	0.2	0.1	0.1	0.2

NOTE: SD = standard deviation.

Results of Models

We report estimated regression coefficients for each model in Appendix G. We also fitted models including linear rather than logged RAND intra-service time as a predictor, which produced similar results (results not reported). We prefer the log-transformed RAND time estimates because they result in fewer extreme predicted values.

We used two approaches to describe how well our models predicted CMS time file values. First, we calculated the fraction of variance explained by each model using the following approach, where y_i is the actual time file value of the dependent variable in question for each i procedure, $\bar{y}$ is the mean of actual time file values across all procedures, and $\hat{y}_i$ is the predicted value from the fitted model for procedure i:

$$Fraction\ of\ variance = 1 - \frac{\sum[(y_i - \hat{y}_i)^2]}{\sum[(y_i - \bar{y})^2]}$$

The fraction of variance describes how well the model accounts for variation in the dependent variable.[25] Second, we calculated the root mean squared error for each model. The root mean squared error is the sum of the squared residuals across all procedures used to fit the model. We normalized residuals (i.e., the differences between predicted and actual CMS time file values) by dividing each residual by the mean of the dependent variable in each model. The root mean squared error describes how far predicted values are "off" from the time file values, on average. Both the fraction of variance and the mixed effect root mean squared errors are reported in Table 5.5. Overall, our mixed-effects model accounted for between 51 and 63 percent of the variation across adjusted time file values. The relatively high values suggest significant agreement.

[25] In the case of ordinary least squares regression, the fraction of variance is equivalent to R^2. However, in the case of generalized linear models (GLMs), it lacks many of the properties of R^2. Notably, the GLMs may include non-constant error variances so that the predictions for some groups of observations are expected, on average, to be farther from the observed values than for other groups of observations. Hence the fraction of variance should not be interpreted as a "goodness-of-fit" measure in this context in the way that R^2 commonly is.

Table 5.5. Model Fraction of Variance Explained and Root Mean Squared Error

Category	Fraction of Variance Explained	Root Mean Squared Error, Mixed Effect Models, Expressed as a Proportion of the Mean Actual Time File Values
Pre-service evaluation	0.63	41.9%
Pre-service positioning	0.61	41.0%
Pre-service scrub	0.51	31.0%
Immediate post-service	0.52	46.7%

Key Predictors of Pre- and Post-Service Time

The RAND intra-service times were usually the most important predictors of each pre-service and immediate post-service time, except for pre-service scrub time. We found that a 10-percent increase in updated intra-service time was associated with a 4.3-percent increase in pre-service evaluation time, a 2.5-percent increase in pre-service positioning time, and a 4.3-percent increase in immediate post-service time.

Procedures with a 90-day global period were associated with 13 percent longer pre-service evaluation times, 32 percent longer pre-service positioning times, and 4 percent longer immediate post-service times compared to procedures without a global period. Several of the CCS level 1 coefficients were statistically significant. While there were few clear patterns, the coefficients for some CCS level 1 categories were significant in both the pre-service evaluation and immediate post-service models.

The patient complexity measures included in the pre-service evaluation and immediate post-service models (including measures of length of stay, complication and mortality rates, etc.) were not consistently statistically significant across models, although some individual coefficients were significant. Median ICU days, median patient age, and median comorbidity count were not significant in either the pre-service evaluation or the immediate post-service models. Procedures with longer median lengths of stay had shorter immediate post-service times, holding other factors constant.

The American College of Surgeons distinguishes between different pre-service time "packages" for laparoscopic and nonlaparoscopic procedures and for thoracic and nonthoracic surgeries (Mabry and Bluff, 2000). We found that laparoscopic procedures have, holding other factors constant, significantly shorter pre-service scrub times, shorter immediate post-operative times, and longer pre-service positioning times compared to nonlaparoscopic procedures. Thoracic procedures have longer pre-service positioning times than nonthoracic procedures.

Comparing Time File and Predicted Values

The predicted time values in each pre-service and post-service component are, on average, slightly shorter than CMS estimates (Table 5.6, Figure 5.1). This is likely due to the log-transformation of the dependent variable. Median time file and predicted values were relatively similar for pre-service evaluation and immediate post-service.

The analyses of pre-service and immediate post-service times presented in this section also apply to pre-service and immediate post-service work because the conversion from time to work involves a constant intensity for each component. Table 5.6 reports RVUs associated with the mean time file and predicted times for reference (only among procedures with reported times).

Table 5.6. Comparison of CMS and RAND Predicted Values for Pre-Service and Immediate Post-Service Components

Work Category	Estimate	Time Mean	Time SD	Time Min	Time P25	Time P50	Time P75	Time Max	Work Mean
Pre-service evaluation	CMS	33.7	24.1	1.0	17.0	27.0	43.0	225.3	0.75
Pre-service evaluation	Typical POS	33.6	19.1	4.1	20.1	28.9	43.9	155.0	0.75
Pre-service evaluation	Across POS	33.6	19.3	4.1	20.0	28.8	43.8	155.6	0.75
Pre-service positioning	CMS	10.0	6.9	1.0	5.0	10.0	15.0	45.0	0.22
Pre-service positioning	Typical POS	10.0	5.3	1.0	5.1	10.2	14.1	31.1	0.22
Pre-service positioning	Across POS	10.0	5.4	1.0	5.0	10.1	14.1	31.1	0.22
Pre-service scrub	CMS	16.2	7.2	1.0	10.0	15.0	25.0	30.0	0.13
Pre-service scrub	Typical POS	16.3	6.9	5.0	15.0	15.0	25.0	25.0	0.13
Pre-service scrub	Across POS	16.3	6.9	5.0	15.0	15.0	25.0	25.0	0.13
Immediate post-service	CMS	25.6	27.5	2.0	15.0	20.0	30.0	401.0	0.57
Immediate post-service	Typical POS	25.3	16.4	3.5	15.9	21.8	30.1	168.5	0.57
Immediate post-service	Across POS	25.3	16.5	3.6	15.8	21.8	30.1	165.8	0.57

Figure 5.1. Time File Versus Fitted (Typical POS) Times

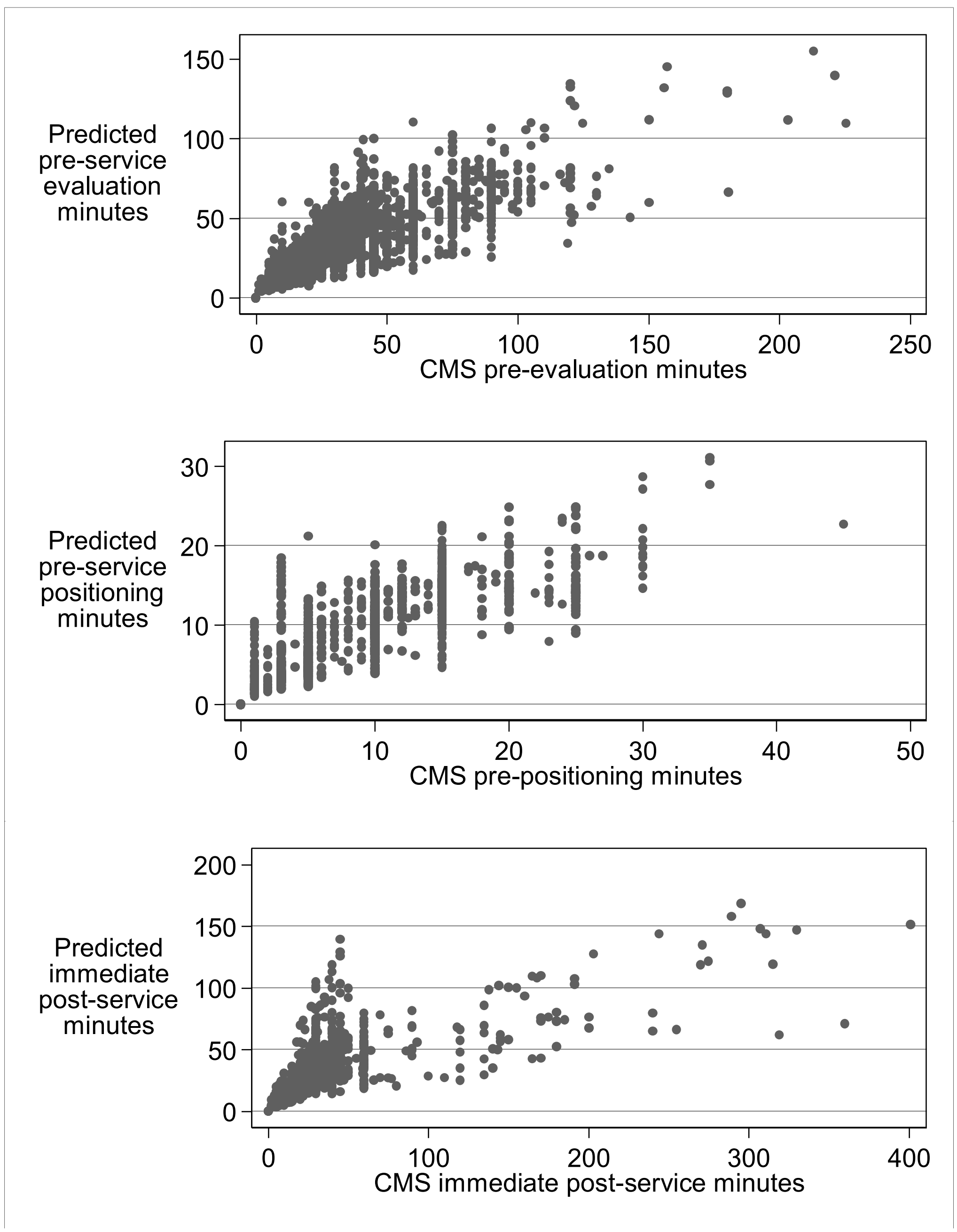


Table 5.7 compares the mean predicted pre-service and immediate post-service work to the mean CMS value. The unweighted means are very similar. Across all procedures, predicted times are associated with 12.9 percent more RVUs (39.7 million RVUs versus 35.3 million RVUs) than the time file values once the changes in pre-service and immediate post-service time and work are weighted by 2012 Medicare utilization volume. This result is driven by longer

predicted times relative to time file times for high-volume codes, and shorter predicted times relative to the time file for low-volume codes.

Table 5.7. Pre-Service and Immediate Post-Service Work, Mean Values at Different Analytic Steps

	CMS Time File	Predicted Value
Unweighted	1.467	1.458
% change from CMS	-	–0.62%
Weighted by 2012 Medicare volume	0.620	0.699
% change	-	12.86%

Table 5.8 reports time file and predicted pre-service and immediate post-service times for 20 high-volume procedure codes. The differences between time file and predicted pre-service and immediate post-service times for specific procedure codes are in some cases extremely large—for example, a near-doubling of pre-service and immediate post-service time associated with CPT 66984, cataract surgery with intraocular lens (1 stage).

Table 5.8. Procedure-Level Comparisons of CMS Time File and Predicted Values for Pre-Service and Immediate Post-Service Work RVUs, Typical POS

Code	Description	Pre-Service Evaluation TF	Pre-Service Evaluation Pred	Pre-Service Positioning TF	Pre-Service Positioning Pred	Pre-Service Scrub TF	Pre-Service Scrub Pred	Post-Service TF	Post-Service Pred	Total RVUs TF	Total RVUs Pred	Total RVUs %Δ
13132	Complex laceration repair	10	15.6	1	4.9	5	5	15	13.4	0.62	0.80	28%
17311	Mohs procedure of skin	14	18.7	1	2.1	5	5	8	13.1	0.56	0.80	44%
20610	Drain or inject a bursa or joint	5	9.1	1	2.2	5	5	5	6.9	0.29	0.45	56%
27245	Treat thigh fracture	40	36.2	30	17.3	20	25	30	23.0	2.40	1.92	–20%
27447	Total knee arthroplasty	40	44.4	15	13.0	20	25	25	26.6	1.95	2.08	7%
31231	Diagnostic nasal endoscopy	5	5.8	1	2.6	5	5	3	5.6	0.24	0.36	47%
33533	CABG arterial single	60	59.3	15	14.8	20	25	40	40.9	2.74	2.78	1%
35301	Rechanneling of artery	40	52.3	15	12.6	20	15	30	38.0	2.07	2.43	17%
43239	Esophagogastroduodeno scopy (EGD) with biopsy	19	16.9	3	3.1	5	5	12	11.1	0.80	0.74	–8%
44120	Removal of small intestine	30	53.8	15	14.7	15	15	30	31.7	1.80	2.37	31%
45380	Colonoscopy with biopsy	45	24.6	0	0.0	0	0	22	17.2	1.50	0.94	–38%
47562	Laparoscopic cholecystectomy	40	41.1	10	10.6	15	5	25	22.0	1.80	1.69	–6%
52000	Cystoscopy	10	12.7	2	6.9	5	15	10	10.8	0.53	0.80	50%
52601	Prostatectomy (TURP)	35	41.8	10	8.8	15	15	40	33.3	2.03	2.00	–1%
62311	Inject spine lumbar/sacral	10	11.6	5	4.8	5	5	10	9.9	0.60	0.63	5%
63047	Remove spinal lamina (lumbar)	40	53.3	15	11.7	20	25	30	32.0	2.07	2.38	15%
64450	Digital nerve block	10	9.3	0	0.0	0	0	5	7.1	0.34	0.37	10%

Code	Description	Pre-Service Evaluation TF	Pre-Service Evaluation Pred	Pre-Service Positioning TF	Pre-Service Positioning Pred	Pre-Service Scrub TF	Pre-Service Scrub Pred	Post-Service TF	Post-Service Pred	Total RVUs TF	Total RVUs Pred	Total RVUs %Δ
66984	Cataract surgery with intraocular lens (1 stage)	16	20.5	1	2.9	5	5	7	21.9	0.58	1.06	83%
67228	Treatment of retinal lesion	24	22.1	12	6.7	12	15	12	15.1	1.17	1.11	–6%
93458	Left heart artery and ventricle angiography	40	37.2	3	3.0	5	5	30	27.3	1.68	1.55	–7%

NOTES: Pred = predicted values; TF = time file.

Typical Versus All POS

We estimated separate typical and all POS models. The two main differences across these models are (1) different RAND intra-service time estimates and (2) different patient complexity independent variables (in the pre-service evaluation and immediate post-service evaluation models only). The differences between predictions from the two sets of models were slight on average (see Table 5.6 above), although the difference at the procedure level is sometimes large (see Table 5.10 below).

Model Predictions in Terms of RVUs

We calculated pre-service and immediate post-service RVUs by multiplying the predicted minutes from each model by the constant intensities associated with each pre-service and immediate post-service component and then summing the results. Table 5.9 reports the volume-weighted average pre-service and immediate post-service RVUs across all procedures (top row) and within groups of codes. Table 5.9 also reports the total RVUs resulting from the typical POS and all POS prediction models. Model predictions were higher for procedures with the shortest intra-service times (by 28 percent on average for the typical POS). Model predictions were 39 percent higher than CMS time file values for procedures most often performed in the emergency department setting and were 27 percent higher than time file values for procedures most often performed in the physician office setting. Model predictions were much higher on average than time file values for procedures in some body systems (e.g., eye and urinary system) and were lower than time file values for procedures in other body systems (e.g., ear and male genital organs).

On average, the predicted values are higher than CMS values for relatively short, outpatient procedures. This drives the weighted results because many of the highest volume procedures fit this description. The unweighted results are reported in Appendix H. Table 5.10 compares CMS versus predicted pre-service and immediate post-service work for a select set of procedures.

Table 5.9. Percentage Difference Between CMS Volume-Weighted Mean Estimates for Pre-Service and Immediate Post-Service Work and RAND Estimates, by Procedure Category

	CMS Work Estimate Weighted Mean RVUs	Typical POS Predicted Work % Difference	All POS Predicted Work % Difference
Total	0.72	11.48	10.99
CMS intra-service time categories			
0 to 30 minutes	0.44	27.82	27.26
31 to 70 minutes	1.11	–0.80	–1.58
71 to 120 minutes	1.62	0.41	0.59
Over 120 minutes	2.38	4.20	4.19
Global period			
0	0.60	13.46	12.72
10	0.60	5.17	5.19
90	1.35	9.40	9.24
Not applicable	0.28	20.71	19.11
Typical place of service			
ED	0.39	38.82	38.87
Inpatient	1.34	2.50	1.31
Office	0.41	27.30	27.41
Ambulatory facility (outpatient or ASC)	0.91	6.43	5.98
Risk category			
Office-based	0.37	29.29	29.10
ASC	0.89	7.55	7.11
Hospital outpatient	1.45	–1.30	–3.17
Inpatient only	2.27	3.49	3.19
Body system grouping			
Nervous system	0.69	4.39	2.46
Endocrine system	1.78	11.35	11.81
Eye	0.51	53.92	53.55
Ear	0.60	–9.21	–9.95
Nose, mouth, and pharynx	0.51	19.34	17.59
Respiratory system	0.77	–1.72	–2.43
Cardiovascular system	1.20	5.29	3.77
Hemic and lymphatic system	1.43	1.81	1.93
Digestive system	1.05	–1.84	–2.36
Urinary system	0.66	21.80	22.32
Male genital organs	1.35	–7.14	–7.45
Female genital organs	1.01	1.98	1.55

	CMS Work Estimate Weighted Mean RVUs	Typical POS Predicted Work % Difference	All POS Predicted Work % Difference
Musculoskeletal system	0.61	17.39	17.53
Integumentary system	0.60	10.67	11.30
Miscellaneous services	0.34	18.51	19.15
Number of annual Medicare procedures			
Less than 1,000	1.63	–0.85	–0.70
1,000 to 9,999	1.37	–1.92	–2.18
10,000 to 99,999	0.95	3.43	3.44
100,000 or more	0.55	21.06	20.14

Table 5.10. Percentage Difference Between CMS Estimates and RAND Estimates for Pre-Service and Immediate Post-Service Work, "Top 20" Codes

Code	Description	CMS Estimate Pre-/Post-Work RVUs	Typical POS Percentage Difference (%)	All POS Percentage Difference (%)
13132	Complex laceration repair	0.6	28.4	26.3
17311	Mohs procedure of skin	0.6	44.0	44.8
20610	Drain or inject a bursa or joint	0.3	56.3	57.5
27245	Treat thigh fracture	2.4	–20.3	–20.7
27447	Total knee arthroplasty	2.0	6.6	6.4
31231	Diagnostic nasal endoscopy	0.2	46.6	45.1
33533	CABG arterial single	2.7	1.4	1.1
35301	Rechanneling of artery	2.1	17.5	16.4
43239	Esophagogastroduodenoscopy (EGD) with biopsy	0.8	–8.0	–9.1
44120	Removal of small intestine	1.8	31.3	31.6
45380	Colonoscopy with biopsy	1.5	–37.7	–38.9
47562	Laparoscopic cholecystectomy	1.8	–6.1	0.8
52000	Cystoscopy	0.5	50.4	51.2
52601	Prostatectomy (TURP)	2.0	–1.2	–1.2
62311	Inject spine lumbar/sacral	0.6	4.8	2.0
63047	Remove spinal lamina (lumbar)	2.1	15.0	14.3
64450	Digital nerve block	0.3	9.7	4.3
66984	Cataract surgery with intraocular lens (1 stage)	0.6	82.7	81.3
67228	Treatment of retinal lesion	1.2	–5.6	–6.6
93458	Left heart artery and ventricle angiography	1.7	–7.3	–14.0

Implications for Our Models

Incorporating Results into the Models

We multiplied the predicted times from each of the four models by constant intensity factors to generate predicted work estimates. We use these estimates in in two ways. First, we subtract these estimates from total work (along with estimates of post-operative E&M work, as described in the next chapter) to calculate our derived intra-service work RVUs. Second, in models using the BBM (Models 1 and 2), the estimates from models described in this chapter contribute to our estimate of total work RVUs.

Limitations

There were no external databases with information on pre-service and immediate post-service times that could be used as a gold standard to build our prediction models. One key limitation is that our models are estimated using the current CMS estimates which may or may not reflect actual times.

While not clearly a limitation, we also emphasize the important assumption of constant intensity within each of the pre-service and the immediate post-service components of work. Without this or another assumption, we would not be able to parse total work into individual components, including intra-service work. The constant intensity values themselves are from the original RBRVS studies and have not been recently updated.

6. Post-Operative Evaluation and Management Work

Overview

This chapter describes the post-operative E&M work component and its valuation using the reverse BBM. We describe corrections that we make in the work values derived from the CMS estimates and how this work component can be predicted for both the typical POS and all POS using characteristics of services from the CMS estimates, the revised time estimates, and Medicare administrative data.

Background

Post-operative E&M visits related to surgical procedures are bundled into total work for CPT procedure codes with a 10- or 90-day global period. These E&M visits are not reimbursed separately under RBRVS when they are performed by the same provider that performed the surgical procedure.[26] For procedure codes with 0-day global periods, any post-operative visits on the day of the surgery are bundled. The CMS time file reports E&M visit counts for 13 individual types of E&M visits (Table 6.1), including inpatient E&M, outpatient E&M, critical care E&M, and discharge E&M visits.

Table 6.1. E&M Services Included in the Global Period

CPT Code	Type of Visit	2014 RVUs If Performed Outside Bundled Payment
99204	Ambulatory	2.43
99211	Ambulatory	0.18
99212	Ambulatory	0.48
99213	Ambulatory	0.97
99214	Ambulatory	1.50
99215	Ambulatory	2.11
99231	Inpatient	0.76
99232	Inpatient	1.39
99233	Inpatient	2.00
99238	Discharge	1.28
99239	Discharge	1.90

[26] Providers use modifier -54 to indicate that they performed the surgical procedure and inpatient E&M visits but not other E&M visits, or modifier -55 to indicate that they performed only outpatient E&M services.

CPT Code	Type of Visit	2014 RVUs If Performed Outside Bundled Payment
99291	Critical care	4.50
99292	Critical care	2.25

Data and Methods for Prediction Models

Providers need not bill for post-operative E&M visits to receive payment for these services because the work associated with the visits is included in the bundled payment for the surgical procedure. As a result, Medicare claims data do not indicate when these post-operative E&M visits are provided or not provided to patients. Ideally, we would use chart or other medical record data to observe the number and type of post-operative E&M visits that are associated with each type of a surgical procedure. Unfortunately, we did not identify any external data—either from Medicare or another source—that describe how often post-operative E&M visits occur.

The CMS time file lists the number of E&M visits by code associated with surgical procedures. We use these visit counts as an input in our models due to the absence of data from external sources. Table 6.2 summarizes the number of post-operative E&M visits for procedure codes with no global period, a 0-day global period, a 10-day global period, and a 90-day global period. Table 6.3 summarizes the number of post-operative visits by typical POS.

Table 6.2. E&M Visits by Type and Global Period, CMS Time File

	0-Day Global Period	0-Day Global Period	10-Day Global Period	10-Day Global Period	90-Day Global Period	90-Day Global Period
E&M Category	**%>0**	**Mean #**	**%>0**	**Mean #**	**%>0**	**Mean #**
Inpatient	1.1	1	4.9	1.8	55.4	4.0
Outpatient	0	-	96.1	1.1	99.9	3.3
Critical care	0	-	0	-	8.0	1.7
Discharge	2.0	0.5	27.4	0.6	82.0	0.9

Table 6.3. E&M Visits by Type and Typical POS

	Inpatient Hospital	Inpatient Hospital	Outpatient Hospital	Outpatient Hospital	Office	Office
E&M Category	**Percentage of Procedures with Visits**	**Mean Visits Procedures with Visits**	**Percentage of Procedures with Visits**	**Mean Visits Procedures with Visits**	**Percentage of Procedures with Visits**	**Mean Visits Procedures with Visits**
Inpatient	78.6	4.4	19.1	1.6	1.3	1.4

	Inpatient Hospital	Inpatient Hospital	Outpatient Hospital	Outpatient Hospital	Office	Office
E&M Category	**Percentage of Procedures with Visits**	**Mean Visits Procedures with Visits**	**Percentage of Procedures with Visits**	**Mean Visits Procedures with Visits**	**Percentage of Procedures with Visits**	**Mean Visits Procedures with Visits**
Outpatient	85.9	3.1	79.8	3.0	62.0	2.0
Critical care	13.1	1.7	0	-	0.2	1.0
Discharge	83.9	1.0	64.7	0.7	7.2	0.6

Most procedures with a 90-day global period have outpatient and discharge post-operative E&M visits, and about half of procedures with a 90-day global period have inpatient post-operative E&M visits. Procedures with a 10-day global period almost always have outpatient post-operative E&M visits. Payments for surgical procedure codes were associated with up to 49.4 million bundled Medicare post-operative E&M visits in 2012. These post-operative E&M visits represented 43.8 million RVUs and payments of $1.57 billion that were bundled into payments for surgical procedures. A few procedures with a 0-day global period have post-operative inpatient (0.7 percent) and discharge (2.2 percent) visits reported in the time file.

Rationale for Making Corrections to Post-Operative E&M Visits

Tables 6.2 and 6.3 point to several potential inconsistencies in the CMS time file. We address these inconsistencies through a series of corrections described below. We implemented one set of corrections for the typical POS models and another set of corrections for the all POS models.

First, some surgical procedures with a 0-day global period have bundled post-operative E&M visits. We assumed that these were in error. For example, they may have been retained on the time file after the procedure code transitioned to 0-day global status. We restricted post-operative E&M visits to no more than one visit for services with a 0-day global code and to zero for procedures that are not subject to the global period policy in both the typical and all POS models.

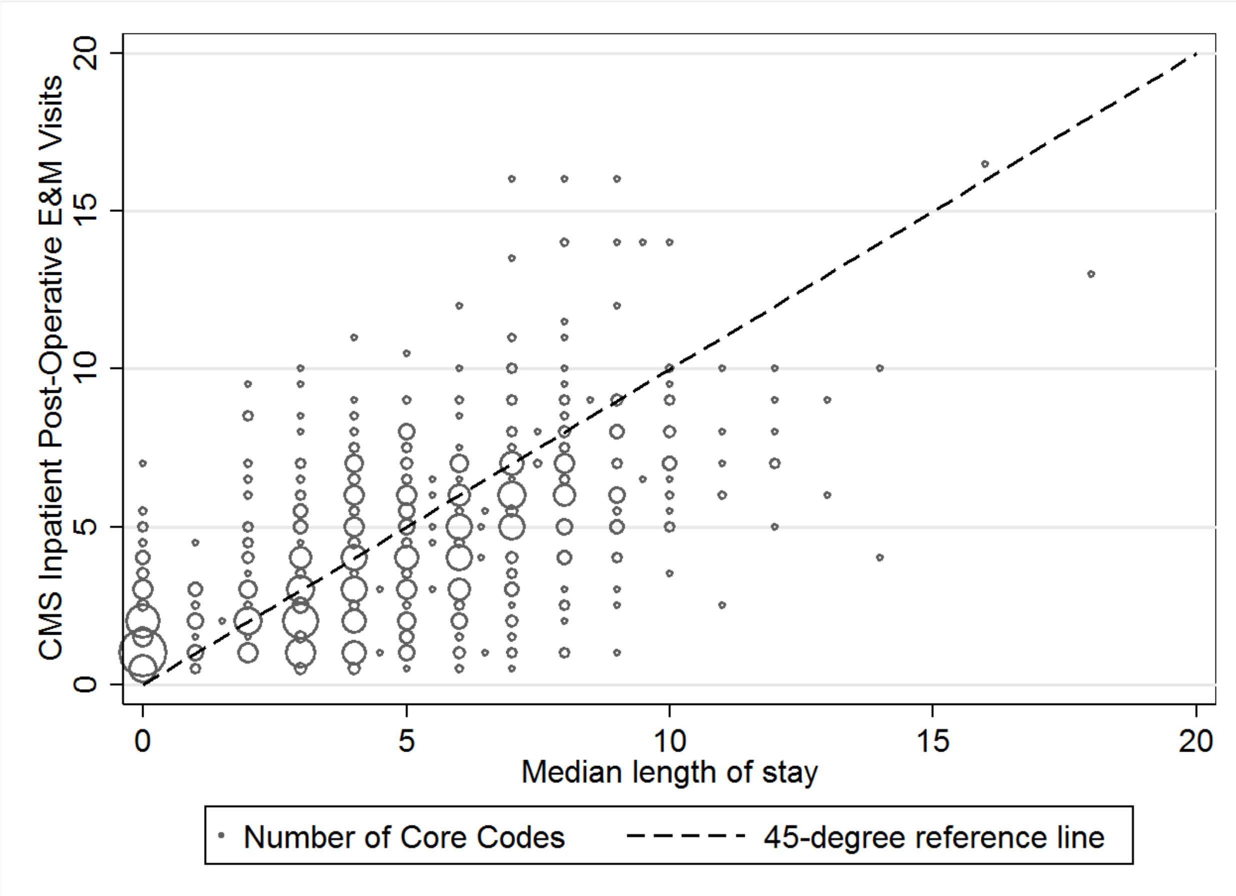
Second, procedures that are not typically performed in the inpatient setting are sometimes assigned inpatient post-operative E&M visits. Because the RUC survey focuses on the typical POS, we assume that there should not be any inpatient post-operative E&M visits listed for procedures most commonly performed in a non-inpatient setting. In the typical models, we restricted inpatient post-operative E&M visits to zero for these procedures. For the all POS models, we reduced the number of inpatient visits in proportion to the share of procedure volume that occurred outside the inpatient setting. For example, for a procedure performed in the inpatient hospital setting 75 percent of the time, performed in the outpatient hospital setting 25 percent of the time, and with 8 total inpatient visits, the corrected inpatient visit count is $(1 - 0.25) \times 8$ visits = 6 visits.

For the all POS models, we also added inpatient E&M visits for procedures that are sometimes but not typically performed in the inpatient setting. We regressed the number of inpatient visits on average length of stay and used the coefficient estimate on length of stay (0.517) multiplied by the median length of stay in the inpatient setting to determine the number of additional inpatient E&M visits. We applied the additional inpatient E&M visits only to the fraction of procedure volume performed in the inpatient setting.

In both the typical and all POS models, we restricted the number of post-operative E&M discharge services to 0.5 for surgical procedures most often performed in the outpatient hospital setting and to 0 for surgical procedures most often performed in an office or other non-hospital settings. These restrictions were applied proportionally in the all POS models. We also added discharge visits in the all POS model for procedures that are not typically performed in the inpatient hospital or outpatient hospital setting but have some volume in one or both of these settings. We added 1 discharge visit times the proportion of volume performed in the inpatient hospital setting and 0.5 discharge visit times the proportion of volume performed in the outpatient hospital setting.

Finally, we made adjustments to align the number of inpatient visits to typical length of stay included in the global period. The post-operative days include the day of the surgery through discharge (or the end of the global period, if earlier). We compared the number of inpatient post-operative E&M visits to the typical (median) post-surgical inpatient length of stay. We calculated median post-surgical length of stay using the date of surgery reported on the physician's bill and the length of stay reported in Medicare data for services provided in the inpatient setting. Figure 6.1 plots typical length of stay against the number of inpatient post-operative E&M visits, with marker size proportional to the number of procedure codes at each point in the plot. Procedure codes plotted above and to the left of the line in Figure 6.1 have more than one inpatient post-operative E&M visit per day. Because Medicare payment rules impose a limit of one inpatient post-operative E&M visit per patient, per provider, and per day, we reduced the number of inpatient visits for procedures to the median length of stay. Procedures above and to the left of the 45-degree reference line in Figure 6.1 were shifted to the reference line.

Figure 6.1. Median Length of Stay Included in the Global Period Versus Number of CMS-Reported Inpatient E&M Visits



After these adjustments, the corrected CMS counts of post-operative E&M visits conform to three rules:

1. Procedures with no global periods are not associated with post-operative E&M visits, and procedures with 0-day global periods are associated with no more than one visit (assumed to occur on the day of surgery).
2. Non-inpatient procedures (or the share of non-inpatient volume in the all POS model) are not associated with inpatient post-operative E&M visits and have at most 0.5 E&M discharge services.
3. Inpatient procedures are associated with a maximum number of inpatient post-operative E&M visits equal to the median inpatient length of stay included in the global period.

While each change to the CMS visit counts affects post-operative work, the impact of a one-visit reduction on work is not always clear because the individual post-operative E&M visits are associated with different RVUs (see Table 6.1). As a result, we calculated a set of post-operative work adjustment factors equal to the average RVUs in each of the visit categories (inpatient, outpatient, critical care, and discharge) for each procedure code. This involved two steps. First, using the time file visit counts and the RVUs associated with each post-operative E&M service, we calculated post-operative work in each of the four post-operative visit categories. Second, we divided work by the number of visits to calculate CPT code-specific RVUs per visit for each category. When post-operative E&M visit counts are adjusted, we add or subtract this adjustment factor from post-operative work.

We considered two options for implementing changes in post-operative E&M work. Under the first option, we reduce total work by the same amount as the post-operative work adjustment. This approach is compatible with the view that any overestimates of post-operative E&M work are also manifest as overestimates of total work. Under the second option, we do not reduce total work. Practically, this means that a reduction in post-operative E&M work results in an increase in intra-service work when applying the BBM. We chose to use the second approach. As discussed earlier in the report, some believe that total work is more accurately estimated than individual work components, and the second approach is more consistent with this theory.

Prediction Models for Post-Operative E&M Visits

Prediction models can identify procedures with more or fewer CMS-reported post-operative E&M visits—and therefore work—than is expected. Prediction models can describe the likely impacts of changes in patient populations or procedure characteristics (such as intra-service time) on the expected number of necessary post-operative visits.

We fit separate CPT code-level Poisson count models predicting the number of post-operative E&M visits in three of the four categories (inpatient, outpatient, and critical care). We fit count models with visit count dependent variables (as opposed to linear models with dependent variables measured in terms of RVUs) because the underlying CMS estimates report counts. Coefficients for each model are estimated using data only from those procedures with at least a fraction of a visit in the given category (after the corrections described above). We subtract one from each dependent variable so that the dependent variable distributions include zero and reflect the number of visits in each category in excess of one.

Independent variables for each model included the log-transformed RAND time estimates, procedure characteristics (including CCS level 1 fixed effects, global period fixed effects, and code grouping random effects), and patient population characteristics (including median length of stay, median number of intensive care unit days, median patient age, proportion of patients that are female, median number of comorbidities, and mortality rate). If the code groupings entered the model as fixed effects, it would be redundant to include the CCS level 1 fixed effects. In the random effects model, however, this is a consequential choice. In particular, random effects that are estimated with little external information will yield estimates that are pulled toward the CCS level 1 estimate rather than the global mean, as would be the case if the level 1 fixed effects were excluded. Because we expect substantial heterogeneity between the level 1 grouping factors, we include them in the model.

For discharge E&M visits, we fit a probit model with a dichotomous dependent variable equal to one if there was a discharge service in the CMS time file and a zero otherwise. The discharge model does not include CCS fixed effects. Due to the small number of codes with critical care E&M visits, we estimate a standard Poisson rather than a mixed effects Poisson model, without CCS or code group fixed effects. As in the previous chapter, we fit these four models for the typical POS and for all POS for a total of eight models. For the typical POS models, the RAND intra-service time estimate and patient characteristic covariates are calculated only for the most common POS. For the all POS models, RAND time and patient characteristics are calculated across all POS. See Appendix C for more detail.

Table 6.4 describes the specific covariates in each model. We fit each model in Stata 13.1 MP using mixed-effects Poisson estimation with the mean and variance adaptive Gauss–Hermite quadrature integration method for random effects. We adjusted standard errors to allow for correlated errors at the code grouping level.

Table 6.4. Dependent and Independent Variables Included in the Post-Operative E&M Models

	Typical Place of Service **(Four models total: inpatient, outpatient, critical care, and discharge)**	**All Places of Service** **(Four models total: inpatient, outpatient, critical care, and discharge)**
Estimation sample	Procedure codes with post-operative E&M visits in the category	Procedure codes with post-operative E&M visits in the category
Dependent variables	Adjusted counts of E&M visits in the inpatient, outpatient, and critical care models; dichotomous discharge visit flag for the discharge category	Adjusted counts of E&M visits in the inpatient, outpatient, and critical care models; dichotomous discharge visit flag for the discharge category
Independent variables	• CCS level 1 fixed effects (omitted for discharge model) • Global code fixed effects • Code group random effects (omitted for discharge model) • Median inpatient length of stay in typical POS • Median ICU days in typical POS • % female in typical POS • Median age in typical POS • Median comorbidity count in typical POS • Mortality rate in typical POS	• CCS level 1 fixed effects (omitted for discharge model) • Global code fixed effects • Code group random effects (omitted for discharge model) • Median length of stay for all POS • Median ICU days for all POS • % female in all POS • Median age in all POS • Median comorbidity count in all POS mortality rate in all POS • Mortality rate in typical POS

Findings

Table 6.5 compares mean CMS post-operative E&M visit counts to visit counts after the corrections described above. The adjustments reduced the average number of inpatient and discharge post-operative E&M visits by 10.1 percent and 16.3 percent, respectively.

Table 6.5. Comparison of Post-Operative Visit Counts Before and After Corrections by Global Period

Global Period	CMS Inpatient Count	CMS Outpatient Count	CMS Critical Care Count	CMS Discharge Count	Corrected CMS Inpatient Count	Corrected CMS Outpatient Count	Corrected CMS Critical Care Count	Corrected CMS Discharge Count
0-day	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00
10-day	0.09	1.05	0.00	0.16	0.08	1.05	0.00	0.13
90-day	2.24	3.25	0.14	0.71	2.01	3.25	0.14	0.60
All	1.49	2.27	0.09	0.49	1.34	2.27	0.09	0.41

Table G.2 in Appendix G reports regression coefficients from our mixed-effects Poisson count models for the typical POS, with one model and table for inpatient, outpatient, critical care, and discharge visit counts. Overall, RAND intra-service time, the code group random effects, mortality rate, and the 90-day global period flag predicted additional counts of post-operative E&M visits. The estimated coefficients for the all POS models share the same overall pattern (not reported).

Figures 6.2 through 6.4 plot the corrected CMS time file counts of inpatient, outpatient, and critical care visits, respectively, on the horizontal axes, compared with the predicted model values on the vertical axes. We did not restrict predicted models to integer values. Table 6.6 compares corrected CMS discharge visit counts to model predictions. For example, there were 34 procedures in the corrected CMS estimates that had a 0.5 discharge visit that we predict to have 0 discharge visits. The prediction model added two procedures to the 518 procedures in the corrected CMS estimates with one discharge visit, bringing the total predicted procedure count to 520 for one discharge visit.

Figure 6.2. Corrected Versus Predicted Post-Operative Inpatient Visit Counts

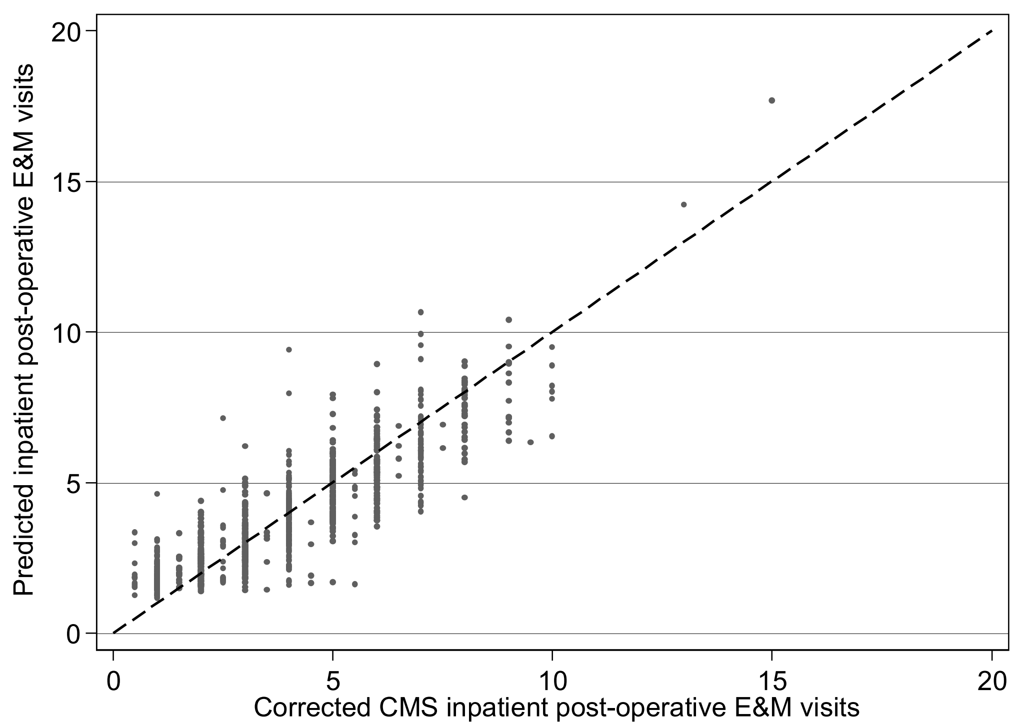


Figure 6.3. Corrected Versus Predicted Post-Operative Outpatient Visit Counts

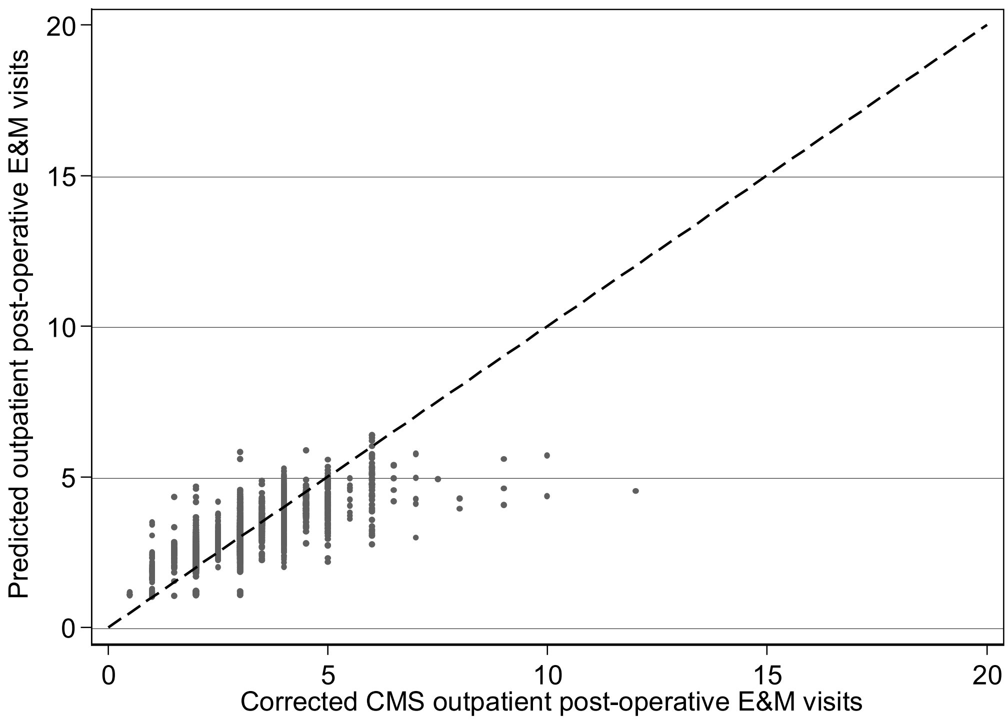


Figure 6.4. Corrected Versus Predicted Post-Operative Critical Care Visit Counts

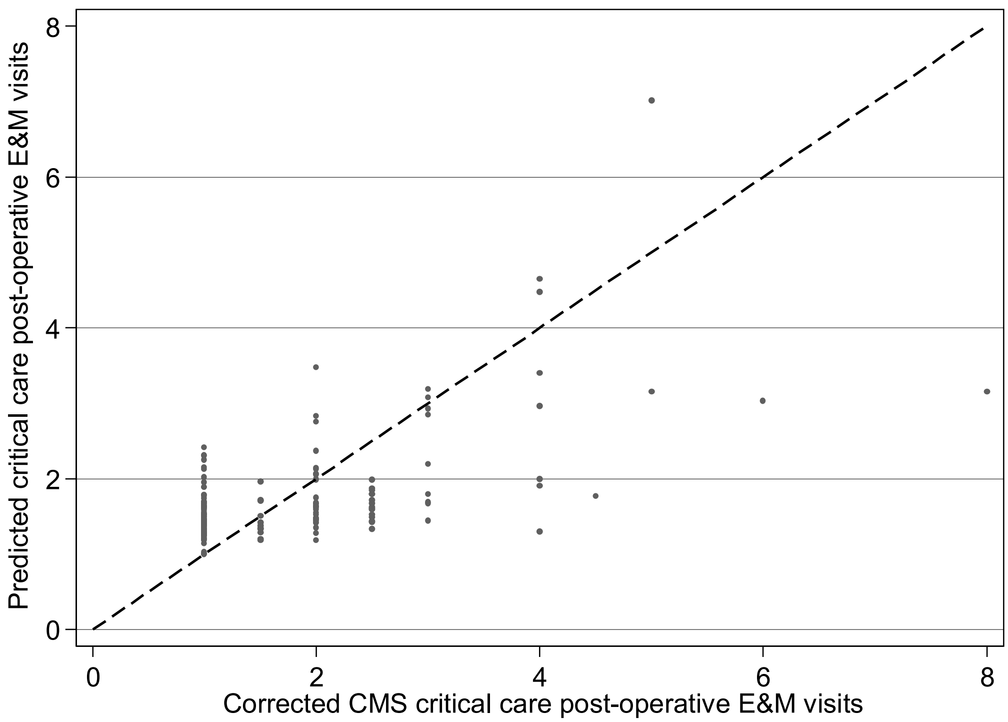


Table 6.6. Procedure Code Counts by Predicted Post-Operative Discharge Visit Count and CMS Corrected Discharge Visit Count Categories

	Corrected CMS Visit: 0.5	Corrected CMS Visit: 1	Any Corrected CMS Visit
Predicted visit: 0	34	0	34
Predicted visit: 0.5	518	2	520
Predicted visit: 1	52	1,003	1,055
Any predicted visit	604	1,005	1,609

In terms of work, post-operative E&M work based on the model predictions is 5.4 percent less than post-operative E&M work based on the time file values (Table 6.7). Table 6.8 reports time file, adjusted time file, and predicted visit counts and work in each post-operative E&M service category for 20 common CPT procedures.

Table 6.7. CMS, Corrected, and Predicted Weighted Mean Visit Counts by Global Period

Global Period	Time File Mean Count	Corrected CMS Mean Count	Predicted Mean Count
10-day	1.30	1.25	1.32
90-day	6.33	6.01	6.01
Total	4.34	4.11	4.12

Table 6.8. Comparison of Mean CMS, CMS Corrected, and Predicted Visit Counts by Category, "Top 20" Procedures

		Inpatient			Outpatient			Critical Care			Discharge		
Code	Description	CMS	Corrected	Predicted	CMS	Corrected	Predicted	CMS	Corrected	Predicted	CMS	Corrected	Predicted
13132	Complex laceration repair	0.0	0.0	0.0	1.0	1.0	1.2	0.0	0.0	0.0	0.0	0.0	0.0
17311	Mohs procedure of skin	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
20610	Drain or inject a bursa or joint	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
27245	Treat thigh fracture	4.0	4.0	2.9	4.0	4.0	3.8	0.0	0.0	0.0	1.0	1.0	1.0
27447	Total knee arthroplasty	3.0	3.0	2.6	3.0	3.0	3.9	0.0	0.0	0.0	1.0	1.0	1.0
31231	Diagnostic nasal endoscopy	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
33533	CABG arterial single	5.0	5.0	5.0	2.0	2.0	2.1	1.0	1.0	1.4	1.0	1.0	1.0
35301	Rechanneling of artery	1.0	1.0	2.0	2.0	2.0	2.0	0.0	0.0	0.0	1.0	1.0	1.0
43239	Esophagogastroduodeno scopy (EGD) with biopsy	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
44120	Removal of small intestine	9.0	8.0	5.8	2.0	2.0	2.4	0.0	0.0	0.0	1.0	1.0	1.0
45380	Colonoscopy with biopsy	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
47562	Laparoscopic cholecystectomy	0.0	0.0	0.0	2.0	2.0	2.2	0.0	0.0	0.0	0.5	0.5	0.5
52000	Cystoscopy	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
52601	Prostatectomy (Turp procedure)	3.0	0.0	0.0	3.0	3.0	2.7	0.0	0.0	0.0	1.0	0.5	0.5
62311	Inject spine lumbar/sacral	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
63047	Remove spinal lamina (lumbar)	2.0	2.0	2.3	3.0	3.0	2.5	0.0	0.0	0.0	1.0	1.0	1.0
64450	Digital nerve block	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

		Inpatient			Outpatient			Critical Care			Discharge		
Code	**Description**	**CMS**	**Corrected**	**Predicted**	**CMS**	**Corrected**	**Predicted**	**CMS**	**Corrected**	**Predicted**	**CMS**	**Corrected**	**Predicted**
66984	Cataract surgery with intraocular lens (1 stage)	0.0	0.0	0.0	4.0	4.0	3.5	0.0	0.0	0.0	0.5	0.0	0.0
67228	Treatment of retinal lesion	0.0	0.0	0.0	3.0	3.0	3.4	0.0	0.0	0.0	0.5	0.0	0.2
93458	Left heart artery and ventricle angiography	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Model Predictions in Terms of Work

We calculated the work associated with predicted post-operative E&M visits by multiplying the number of predicted visits in each of the four categories by a category-specific intensity factor. The predicted post-operative E&M work is the sum of these four individual work values.

$$\begin{aligned}&\textit{Post-operative E\&M RVUs}_{predicted}\\&\quad = \textit{Inpatient RVUs}_{predicted} + \textit{Outpatient RVUs}_{predicted}\\&\quad + \textit{Critical care RVUs}_{predicted} + \textit{Discharge RVUs}_{predicted}\end{aligned}$$

The four post-operative E&M intensity factors—one each for inpatient, outpatient, critical care, and discharge—are procedure-specific factors calculated by dividing the CMS post-operative work in each category by the CMS number of visits in each category. The following equation illustrates our approach for inpatient visits.

$$\textit{Inpatient RVUs}_{predicted} = \textit{Inpatient visits}_{predicted} \times \frac{\textit{CMS Inpatient RVUs}}{\textit{CMS Inpatient visits}}$$

Table 6.9 reports the weighted mean CMS post-operative E&M work, as well as the percentage change in the weighted mean from the CMS values in the typical POS models and the all POS models. Across all codes (and weighted by 2012 Medicare volume), predicted post-operative E&M RVUs are 8.2 percent lower than CMS RVUs on average in the typical POS models and 4.3 percent lower in the all POS models. The average difference between predicted and CMS values is larger for procedures with short intra-service times (15.1 percent lower); procedures with 0-day global periods (81.3 percent lower); and procedures focusing on the nervous system, eye, urinary, female genital, and male genital body systems. Some of these procedure categories—for instance, codes with 0-day global periods—were specifically targeted in our post-operative correction steps. On average, the largest percentage reductions occur for procedures most often performed in ambulatory facilities or offices and for ASC-covered procedures. There are increases in a few CCS level 1 groupings—e.g., ear and the cardiovascular system. Table 6.10 compares procedure-level CMS and predicted values for a set of illustrative procedure codes. Several of these codes had zero post-operative E&M visits in the CMS time file and as a result have zero predicted post-operative RVUs.

Table 6.11 describes the relative contributions of the correction and prediction steps. Panel A compares predicted total post-operative work for all codes to (a) the mean CMS time file value, and (b) the mean CMS value after the correction steps outlined above. Total predicted post-operative work is the sum of four separate predictions for inpatient, outpatient, critical care, and discharge visits. The correction steps reduce mean post-operative work by 10.2 percent in the unweighted results and 8.5 percent in the weighted results. The prediction step offsets this reduction somewhat. Panels B and C report similar patterns for procedures with 90- and 10-day global periods separately.

Table 6.9. Percentage Difference Between CMS Weighted Mean Post-Operative Estimates and RAND Estimates, by Procedure Category

Procedure Category	CMS Estimate Mean Work RVUs	Typical POS Predicted Work % Difference	All POS Predicted Work % Difference
Total	0.90	–8.18	–4.30
CMS intra-service time categories			
0 to 30 minutes	0.31	–15.08	–14.15
31 to 70 minutes	0.86	–5.64	–8.28
71 to 120 minutes	4.00	–7.07	2.53
Over 120 minutes	9.60	–5.23	0.57
Global period			
0	0.00	–81.28	–100.00
10	0.82	–5.82	–8.90
90	4.49	–8.20	–3.53
Not applicable	0.00	N/A	N/A
Typical place of service			
ED	0.15	–3.15	1.46
Inpatient	2.88	–2.21	4.84
Office	0.25	–11.59	–12.07
Ambulatory facility (Outpatient or ASC)	0.98	–14.64	–13.25
Risk category			
Office-based	0.16	–8.34	–8.73
ASC	0.97	–11.51	–11.54
Hospital outpatient	1.23	–7.23	1.24
Inpatient only	9.35	–3.50	6.25
Body system grouping			
Nervous system	0.57	–17.93	–10.80
Endocrine system	2.84	–4.74	4.85
Eye	1.76	–15.55	–14.92
Ear	0.93	5.34	6.04
Nose, mouth, and pharynx	0.28	–10.44	–11.50
Respiratory system	0.41	–5.25	3.81
Cardiovascular system	1.07	6.04	5.40
Hemic and lymphatic system	2.39	–1.94	–3.06
Digestive system	0.70	–3.04	–3.68
Urinary system	0.36	–15.06	–9.15
Male genital organs	1.92	–29.06	–6.43
Female genital organs	1.53	–24.20	–13.08

Procedure Category	CMS Estimate Mean Work RVUs	Typical POS Predicted Work % Difference	All POS Predicted Work % Difference
Musculoskeletal system	1.23	–5.44	3.39
Integumentary system	0.68	–8.58	–7.44
Miscellaneous services	0.01	–13.78	–12.92
Number of annual Medicare procedures			
Less than 1,000	5.22	–10.89	–5.41
1,000 to 9,999	3.12	–8.05	–5.26
10,000 to 99,999	1.37	–6.08	–3.03
100,000 or more	0.40	–10.39	–4.90

Table 6.10. Percentage Difference Between CMS Estimate Weighted Mean Post-Operative Work Estimates and RAND Estimates, "Top 20" Codes

Code	Description	CMS Estimate Mean Work RVUs	Typical POS Predicted Work % Difference	All POS Predicted Work % Difference
13132	Complex laceration repair	0.5	17.6	17.5
17311	Mohs procedure of skin	0.0	0.0	0.0
20610	Drain or inject a bursa or joint	0.0	0.0	0.0
27245	Treat thigh fracture	9.0	–14.5	–1.1
27447	Total knee arthroplasty	7.7	4.8	23.8
31231	Diagnostic nasal endoscopy	0.0	0.0	0.0
33533	CABG arterial single	15.9	12.6	12.7
35301	Rechanneling of artery	4.6	29.0	40.2
43239	Esophagogastroduodenoscopy (EGD) with biopsy	0.0	0.0	0.0
44120	Removal of small intestine	13.9	–26.5	–26.0
45380	Colonoscopy with biopsy	0.0	0.0	0.0
47562	Laparoscopic cholecystectomy	2.1	5.7	5.2
52000	Cystoscopy	0.0	0.0	0.0
52601	Prostatectomy (TURP)	6.6	–57.7	–26.3
62311	Inject spine lumbar/sacral	0.0	0.0	0.0
63047	Remove spinal lamina (lumbar)	6.3	–1.1	14.2
64450	Digital nerve block	0.0	0.0	0.0
66984	Cataract surgery with intraocular lens (1 stage)	3.5	–28.1	–26.6
67228	Treatment of retinal lesion	3.6	–0.8	–5.2

Code	**Description**	**CMS Estimate Mean Work RVUs**	**Typical POS Predicted Work % Difference**	**All POS Predicted Work % Difference**
93458	Left heart artery and ventricle angiography	0.0	0.0	0.0

Table 6.11. Total Post-Operative Work, Mean Values at Different Analytic Steps

A: All Core Codes

	CMS Time File	**After Correction Steps**	**Predicted Value**
Unweighted mean	4.23	3.80	3.81
% difference from CMS estimate	-	–10.2%	–9.8%
Volume-weighted mean	0.90	0.82	0.82
% difference from CMS estimate	-	–8.5%	–8.1%

B: Codes with a 90-Day Global Period

	CMS Time File	**After Correction Steps**	**Predicted Value**
Unweighted mean	6.17	5.54	5.57
% difference from CMS estimate	-	–10.1%	–9.8%
Volume-weighted mean	4.49	4.11	4.12
% difference from CMS estimate	-	–8.5%	–8.2%

C: Codes with a 10-Day Global Period

	CMS Time File	**After Correction Steps**	**Predicted Value**
Unweighted mean	0.93	0.86	0.90
% difference from CMS estimate	-	–7.8%	–3.20%
Volume-weighted mean	0.82	0.78	0.78
% difference from CMS estimate	-	–5.8%	–5.8%

Limitations

We did not identify any useful external databases with information on the number of post-service E&M visits associated with surgical procedures. We could not use Medicare claims data because post-operative E&M visits in the global period are bundled into the payment for surgical procedures and are not reported separately. As a result, we fit prediction models using visit counts from the CMS time file as dependent variables. As discussed in Chapter 1, there is concern that visit counts from the CMS time file may be overestimated. This potential bias is incorporated into our models.

We assumed that changes in post-operative visit counts influence post-operative work in proportion with procedure-specific adjustment factors. These factors were the procedure-specific average RVUs in each post-operative visit category. As a result, the impact of actual revisions in post-operative E&M visit counts may be higher or lower than our model suggests if the corrections target specific visits (i.e., visits with especially large or small RVU values for post-operative E&M visits compared to other visits). Finally, our predicted values accommodate fractional visits. While fractional visit counts occasionally appear on the time file, these are often limited to half or quarter visits.

We limit the number of inpatient hospital visits to the Medicare post-operative median length of stay for the procedure covered by the global period. There may be situations where multiple inpatient visits may occur on the same day. Under Medicare rules, only a single visit would be billed but at a level that reflects the total work performed that day. This would result in a higher average intensity per inpatient visit than is reflected in the average intensity used in the model.

There has been a significant increase in Medicare hospital observation days. Hospital observation days involve a patient spending up to 48 hours under observation as an outpatient. There may or may not be a subsequent inpatient admission. This may affect the typical POS and the inpatient median length of stay (i.e., if following outpatient surgery a patient is retained for observation and subsequently admitted as an inpatient). To some extent, observation visits may be substituting for inpatient visits. We discuss this issue further in Chapter 9.

Implications for Our Models

The post-operative E&M analyses implement several corrections to CMS visit counts and predict new post-operative visit counts and work using multivariate models. The work estimates are important inputs into each of our BBM models.

We corrected post-operative E&M visit counts and post-operative E&M work because current CMS counts may have errors. As a result, the prediction models are estimated using more accurate and consistent post-operative E&M visit values. For the most part, the corrections to post-operative E&M work were slight. In Chapter 7, we discuss how these corrections affect the derivation of intra-service work through the reverse BBM.

We also chose to model post-operative E&M work in terms of visit counts rather than RVUs. This decision allows us to adjust post-operative E&M visits (e.g., in response to length of stay restrictions). Count rather than linear models seem to be a more appropriate fit for post-operative E&M visits because the underlying CMS estimates are reported in terms of visit counts rather than RVUs. The challenge with this approach is that different E&M visits are associated with different intensity and work values. We used four separate prediction models—one each for inpatient, outpatient, critical care, and discharge visits—in recognition of heterogeneity across specific E&M codes. In a necessary simplifying assumption, we assumed that the average intensity for a specific procedure code is constant within each of these four categories, so that we can calculate predicted post-operative work by multiplying the predicted number of visits by the average RVUs per visit.

As in the previous chapter, we estimated separate typical and all POS models. The two main differences across these models in terms of explanatory variable values are (1) different RAND intra-service time estimates and (2) different approaches to calculate patient complexity independent variables (in the pre-service evaluation and immediate post-service evaluation models only). Overall, however, the two approaches result in similar corrected and predicted estimates of post-operative E&M work.

Finally, our estimations are based on calendar year (CY) 2014 CMS policies regarding global periods for surgical procedures. In its CY 2015 proposed rule for the Medicare physician fee schedule, CMS proposed to transition to 0-day global periods beginning in CY 2017 for the current 10-day global periods and in CY 2018 for the 90-day global periods. We explore in Chapter 9 what modifications would be needed if our models were used to validate physician work for 0-day global periods.

7. Intra-Service Work and Intra-Service Intensity (IWPUT)

Overview

This chapter describes the different models of intra-service work and intra-service intensity (IWPUT). First, we give background on the distinction between intra-service work and IWPUT and different ways they can be generated. Then, we provide the rationale for our predictive models. We also discuss why certain procedure characteristics may be correlated with intra-service work and intensity and show the actual correlations. We compare the intra-service work and IWPUT estimates derived from the CMS time file to our estimates.[27] Based on these findings, we discuss the rationale for three different models to predict intra-service work and compare the RAND model results with those in the time file. We end with a discussion of the implications for our approaches to estimating total work.

Background on Intra-Service Work and Intra-Service Intensity (IWPUT)

The model developed in the original Harvard study (Hsiao et al., 1988, 1992) conceptualizes physician work as a function of both time and intensity (i.e., work equals the product of time and intensity). Qualitative research conducted by Hsiao et al. (1992) during the early stages of development of the RBRVS found that physician work consisted of the following dimensions: time, technical skill and physical effort, mental effort and clinical judgment, and psychological stress and risk. These non-time dimensions capture the intensity of the service. For example, a service that requires greater technical skill will have a higher intensity and should be compensated at a higher rate than another service with the same amount of time that requires less technical skill. In the same way, a service that involves greater risk (and therefore psychological stress) should be compensated at a higher rate than a routine service.

Because of empirical difficulties in measuring intensity independently, the original Harvard study and the current RUC process instead estimate *work* and *time* independently. Given estimates of both work and time, one can then "back out" estimates of intensity by dividing work by time. For intra-service work, we have only a measure of intra-service time and need to derive the intra-service work and intensity using the reverse BBM. This is because the RUC does not currently collect physicians' direct estimates of intra-service work.

We use the following formula to derive an initial estimate of intra-service work.

[27] We use the term *derived* because they are not based on the CMS time file, but rather use the estimated RAND time and pre- and post-service work estimates described in prior chapters.

$$\begin{aligned} \textit{Intra-service work}_{\textit{RAND derived}} \\ &= \textit{Total work RVUs} - \textit{Pre-service work}_{\textit{predicted}} \\ &- \textit{Immediate post-service work}_{\textit{predicted}} - \textit{Post-operative E\&M work}_{\textit{predicted}} \end{aligned}$$

In calculating this intra-service work we subtract values from the prediction models described in Chapters 5 and 6. Dividing derived intra-service work by RAND intra-service time yields a derived intra-service intensity per unit of time (IWPUT) for each core procedure.

$$IWPUT_{RAND\ derived} = \frac{\textit{Intra-service work}_{\textit{RAND derived}}}{RAND\ intra\text{-}service\ time}$$

We use the term *RAND derived* to distinguish these values of intra-service work and IWPUT from what we would obtain if we calculated intra-service work and IWPUT using values for pre-service work, immediate post-service work, and post-service E&M work from the CMS time file. We can calculate the $\text{IWPUT}_{\textit{CMS derived}}$ in a similar manner also using the reverse BBM.

$$\begin{aligned} \textit{Intra-service work}_{\textit{CMS derived}} \\ &= \textit{Total work RVUs} - \textit{Pre-service work}_{\textit{CMS}} - \textit{Immediate post-service work}_{\textit{CMS}} \\ &- \textit{Post-operative E\&M work}_{\textit{CMS}} \end{aligned}$$

$$IWPUT_{CMS\ derived} = \frac{\textit{Intra-service work}_{\textit{CMS derived}}}{RAND\ intra\text{-}service\ time}$$

Obviously these sets of intra-service work and IWPUT are related, but there are important differences that we discuss later in this chapter.

These different mechanisms to measure intra-service work and IWPUT become important in considering how to model intra-service work and IWPUT. When creating our models, we had to consider some important issues. The first is that we have no independent measure or "gold standard" for IWPUT. While IWPUT can be measured indirectly, interpretation of what is the right IWPUT is a major issue driving the different models we create. In the original development of the RBRVS the average IWPUT for a surgical procedure was 0.057. Whether that average value should be different now is unclear, given the changes in practice patterns that have occurred since the original valuations. A second issue is nonsensical intra-service work and IWPUT values. As detailed in Chapter 1, one criticism of the current system is that for a small fraction of codes, the CMS intra-service work and therefore IWPUT values are negative or so low that they lack face validity. This is something we describe in more depth below and is another key issue that we address in our modeling efforts.

Rationale for Using Procedure Characteristics to Model Intra-Service Work and IWPUT

As noted above, Hsiao et al. (1992) during the early stages of development of the RBRVS found that physician work had the following non-time elements: (1) technical skill and physical effort, (2) mental effort and clinical judgment, and (3) psychological stress and risk. We cannot directly measure these non–time element domains, but we identified code characteristics that theoretically could be correlated with one or more of the domains described by Hsiao et al. (1992) as part of service intensity. Table 7.1 provides an overview of some of these variables and indicates the domains that each might be associated with.

Table 7.1. Procedure Characteristics Potentially Correlated with Intra-Service Intensity by Domains of Intensity

Procedure Characteristics	Technical Skill/ Physical Effort	Mental Effort/ Clinical Judgment	Psychological Stress/Risk
Years of clinical training	X	-	-
Mortality risk	X	X	X
Complication rates	X	X	X
Length of stay	X	X	X
ICU days	X	X	X
Malpractice risk	-	-	X
Patient characteristics	-	X	X
Urgency of decisionmaking	-	-	X

We identified five variables that may be associated with the first domain of intensity: technical skill/physical effort. Procedures performed by physicians with more years of clinical training should, in theory, reflect procedures that require greater technical skill. Procedures with greater risk of mortality and complications, as well as longer length of stay and more ICU days, are likely more complex and require more technical skills.

We believe that five variables could also be associated with greater mental effort and clinical judgment, the second domain. Procedures performed on patients who are on average sicker or procedures with greater risk of complications and mortality are likely procedures that require greater mental effort and judgment. Similarly, patient characteristics, length of stay, and ICU days could be markers of risk of a procedure.

Finally, we added malpractice risk (as measured by relative malpractice premiums paid for by physicians who perform the procedure) and urgency of decisionmaking (as measured by how often procedure is performed in an emergency department with a subsequent hospital admission or on the first day of hospital stay if they were admitted from the emergency department) are potential markers of psychological stress. Details on how we estimated each of the variables are included in Appendix C. Below we present the actual association between these code characteristics, IWPUT, and intra-service work.

Addressing Negative Intra-Service Work and IWPUT

A key consideration is developing intra-service work and IWPUT models is nonsensical values. There is concern that some values in the current CMS file are too low. Because the intra-service work and IWPUT values are derived from total work, these values are dependent on the values for the other work components. For example, an overvaluation of the post-operative E&M visits can lead to an undervaluation of intra-service work.

As points of reference, the $IWPUT_{CMS}$ for most pre-service and immediate post-service work is 0.0224, and for a very complex procedure, such as CPT 32854 (lung transplant with bypass), it is 0.11. We view these as reference points to create an intensity range and examine what fraction of codes is outside that range. In current CMS estimates, 2.7 percent of procedure codes have negative $IWPUT_{CMS,}$ and an additional 7.8 percent of procedure codes have $IWPUT_{CMS}$ between 0 and 0.0224 RVUs/minute. Only 4.8 percent of procedure codes have $IWPUT_{CMS}$ over 0.11.

We are most concerned about nonsensical negative IWPUT (and therefore negative intra-service work values) and about extremely small IWPUT values that do not have face validity relative to other well-known procedures like E&M visits and pre-service work. In our predictive models, we therefore impose floors on intensity and intra-service work. For intra-service intensity, the floor is 0.0224 RVUs/minute. For intra-service work, the floor is the product of 0.0224 RVUs/minute multiplied by the RAND time. We choose 0.0224 RVUs/minute because this is the intensity assigned to pre-service evaluation, pre-service positioning, and immediate post-service work.

The overall impact of this correction step is relatively small. Mean derived intra-service work increases from 5.25 to 5.35 RVUs. Mean $IWPUT_{CMS}$ increases from 0.057 to 0.059 RVUs/minute.

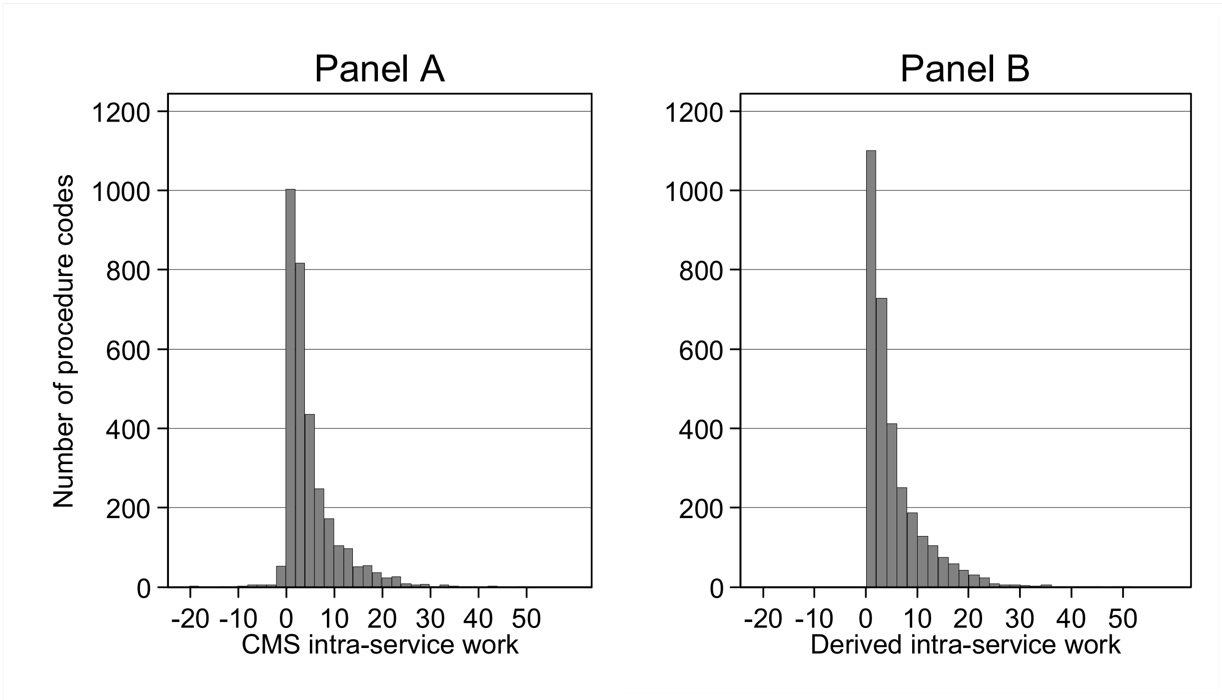
Descriptive Results

We begin with descriptive results showing the distribution of intra-service work and intra-service intensity.

Distribution of Intra-Service Work

Figure 7.1 compares the distribution of intra-service work values derived from the CMS estimates (intra-service $work_{CMS}$) in Panel A to the values derived from our predicted values for the other work components (intra-service $work_{RAND}$) in Panel B. The formulas for each are described above.

Figure 7.1. Distribution of CMS Intra-Service Work (Panel A) and RAND Intra-Service Work (Panel B)



The RAND-derived estimates of intra-service work have a slightly smoother distribution and no negative values, due to the floors that we describe above.

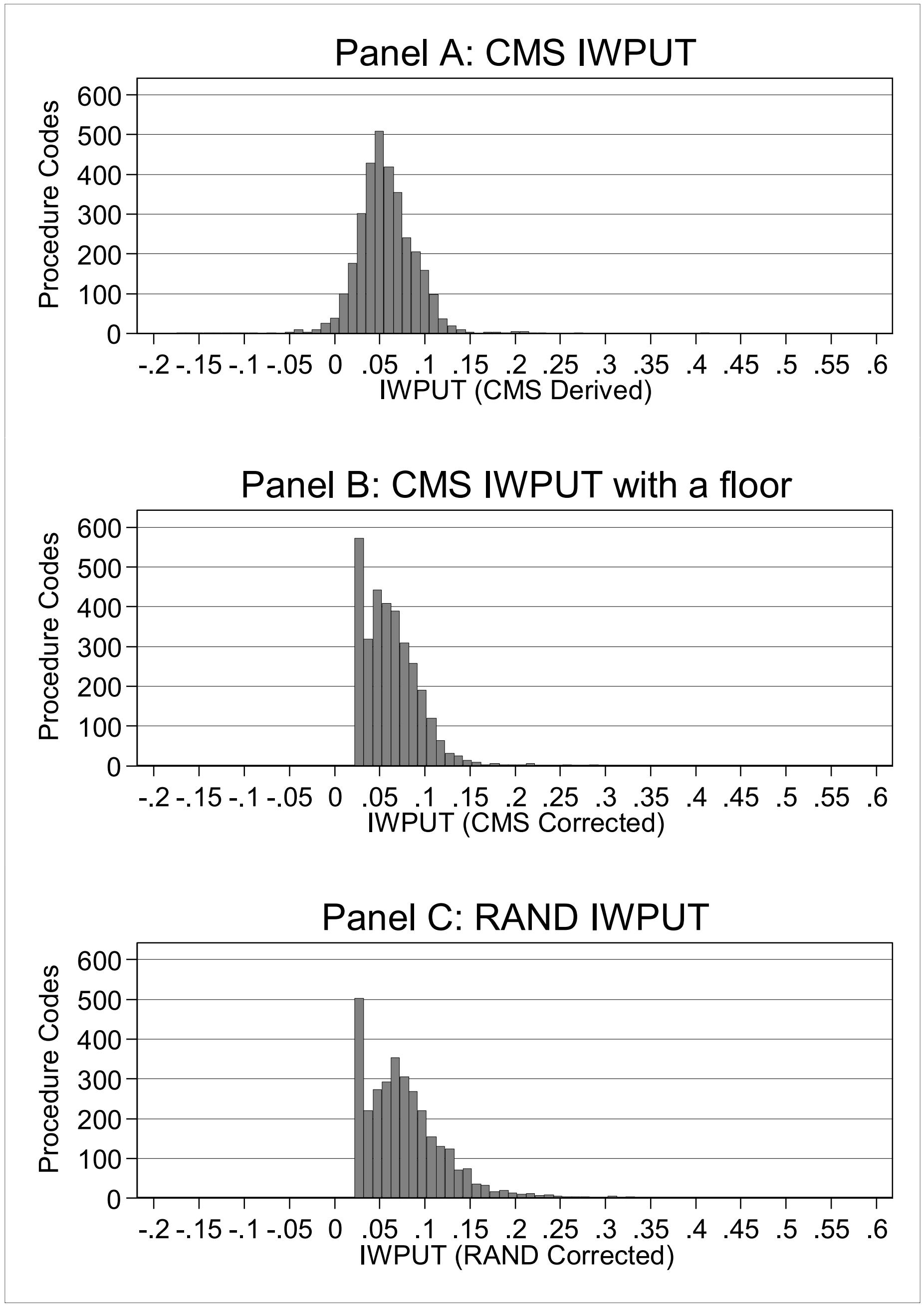
Distribution of IWPUT

In Figure 7.2, we compare $IWPUT_{CMS\ derived}$ (Panel A) with $IWPUT_{CMS\ corrected}$ (Panel B) and $IWPUT_{RAND\ corrected}$ (Panel C). The $IWPUT_{RAND\ corrected}$ reflects the addition of the 0.0224 floor. The IWPUT values are calculated from RAND-derived intra-service work after the floor is imposed but before intra-service work is predicted. We compare them to $IWPUT_{CMS\ derived}$ because they provide background on one of the issues we considered in developing our models (Issue C). Because the RAND intra-service time estimates are systematically lower than the CMS time estimates, the average IWPUT values are much higher if we use RAND time estimates (Panel C) than if we used CMS time estimates with corrections (Panel B) (Table 7.2). One key consideration in developing our models is whether IWPUT should be higher overall (Model 1a) or whether the average new IWPUT values should be similar to the current IWPUT values (Model 1c). As discussed later in this chapter, the values in Panel B are used to predict IWPUT for Model 1c.

Table 7.2. Mean IWPUT Across Different Ways of Measuring IWPUT

IWPUT Measure	Explanation	Mean IWPUT
$\mathbf{IWPUT_{CMS\ derived}}$	Derived from CMS estimates	0.057
$\mathbf{IWPUT_{CMS\ corrected}}$	$IWPUT_{CMS\ derived}$ after corrections for the 0.0224 floor	0.059
$\mathbf{IWPUT_{RAND\ corrected}}$	Calculated from RAND-derived intra-service work with corrections ÷ RAND time	0.077

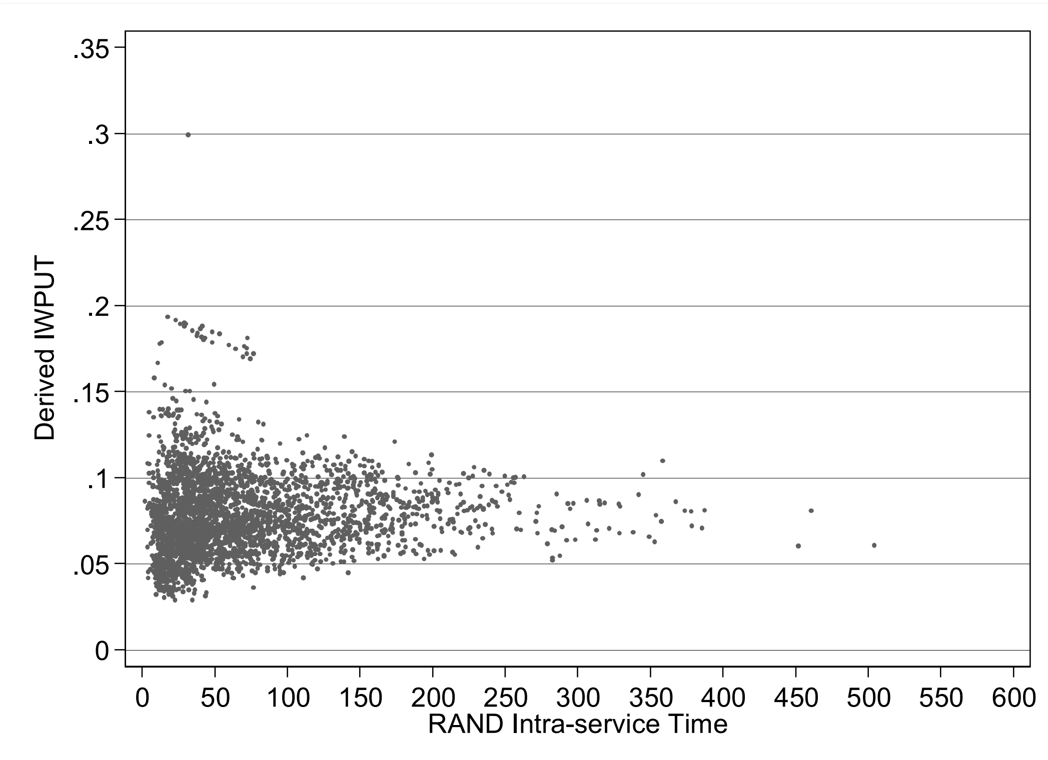
Figure 7.2. Distribution of IWPUT Values



Correlation Between IWPUT, Intra-Service Work, and Procedure Characteristics

The two components of intra-service work are intra-service time and IWPUT. In Figure 7.3, we show the relationship between $IWPUT_{RAND\ derived}$ and RAND intra-service time. Overall, there is no significant correlation between $IWPUT_{RAND\ derived}$ and RAND intra-service time ($p=0.91$). However, when we separately analyze procedure codes with $IWPUT_{RAND\ derived}$ under 0.2, there is a positive and significant correlation between IWPUT and RAND intra-service time (correlation coefficient of 0.09, $p<0.001$).

Figure 7.3. RAND Intra-Service Time Versus Derived IWPUT



Above we highlighted different procedure characteristics that could be correlated with intra-service work and IWPUT. Most of the "proxies for intensity" are positively and significantly correlated with intra-service work, measured either from the CMS estimates or derived RAND estimates (Table 7.3).[28] Length of stay, ICU days, comorbidity count, malpractice index, and the training index are most strongly correlated. Median patient age and the urgency of decision-making index are slightly negatively correlated with intra-service work. In contrast, most of the proxies for intensity have a weak relationship with derived IWPUT. Higher mortality risk, age, comorbidity counts, and ICU days are all correlated with higher IWPUT, but the correlation is often weak. Some correlations are also contrary to expectations. For example, a longer length of stay is associated with a *lower* IWPUT.

---

[28] This is for typical POS; results for all POS are similar (not shown).

This lack of relationship between IWPUT and proxies for intensity is largely explained by the relationship between intra-service time and proxies for intensity. RAND time is correlated with the training index (correlation coefficient 0.44), malpractice risk (0.31), and mortality (0.19). After controlling for time, these proxies have little residual relationship. In some sense, this is consistent with the original work by Hsiao and colleagues where physicians had difficulty distinguishing between time and intensity.

Table 7.3. Correlation Between Code Characteristics and Intra-Service Work and IWPUT

	Median Length of Stay	Median ICU Days	Median Patient Age	Median Comorbidity Count	Mortality Rate	Urgency of Decision-making	Malpractice Risk	Training Index
$IWPUT_{CMS\text{-}C}$	0.05*	0.09*	0.07*	0.11*	0.06*	–0.04*	–0.05*	0.03
$IWPUT_{RAND}$	–0.10*	–0.02	0.07*	0.03	–0.04*	–0.06*	–0.08*	–0.04*
CMS intra-service work	0.45*	0.33*	–0.03*	0.25*	0.14*	–0.05*	0.25*	0.35*
RAND intra-service work	0.46*	0.34*	–0.04*	0.25*	0.16*	–0.06*	0.25*	0.36*

* = p<0.05.

Model Options Related to Intra-Service Work or IWPUT

In creating our models (Figure 7.4), one key consideration was how to model intensity (Issue C). The core question underlying Issue C is the implications of reduced RAND time on IWPUT. In the original development of the RBRVS 20 years ago, the mean IWPUT for surgical procedures was 0.057, and in the current CMS time file mean IWPUT for surgical procedures is 0.056 (Table 7.2). As noted in Chapter 4, the RAND time estimates are shorter on average than the CMS time file estimates. If intra-service work values remain the same and we divide by our reduced RAND time estimates, mean IWPUT will increase on average. Should mean IWPUT remain the same or increase?

One argument for maintaining the mean IWPUT is that it reflects a core level of intensity for surgical procedures. It is notable that the IWPUT in the time file for surgical procedures is almost exactly the same as it was in the original development. It could even be argued that mean IWPUT should be lower. Mortality after surgical procedures has progressively declined, and improved technology and efficiency gains mean that surgical procedures can be done more easily and safely. Together this might imply that physical effort, mental effort, and psychological stress and risk (all core aspects of intensity) have also decreased.

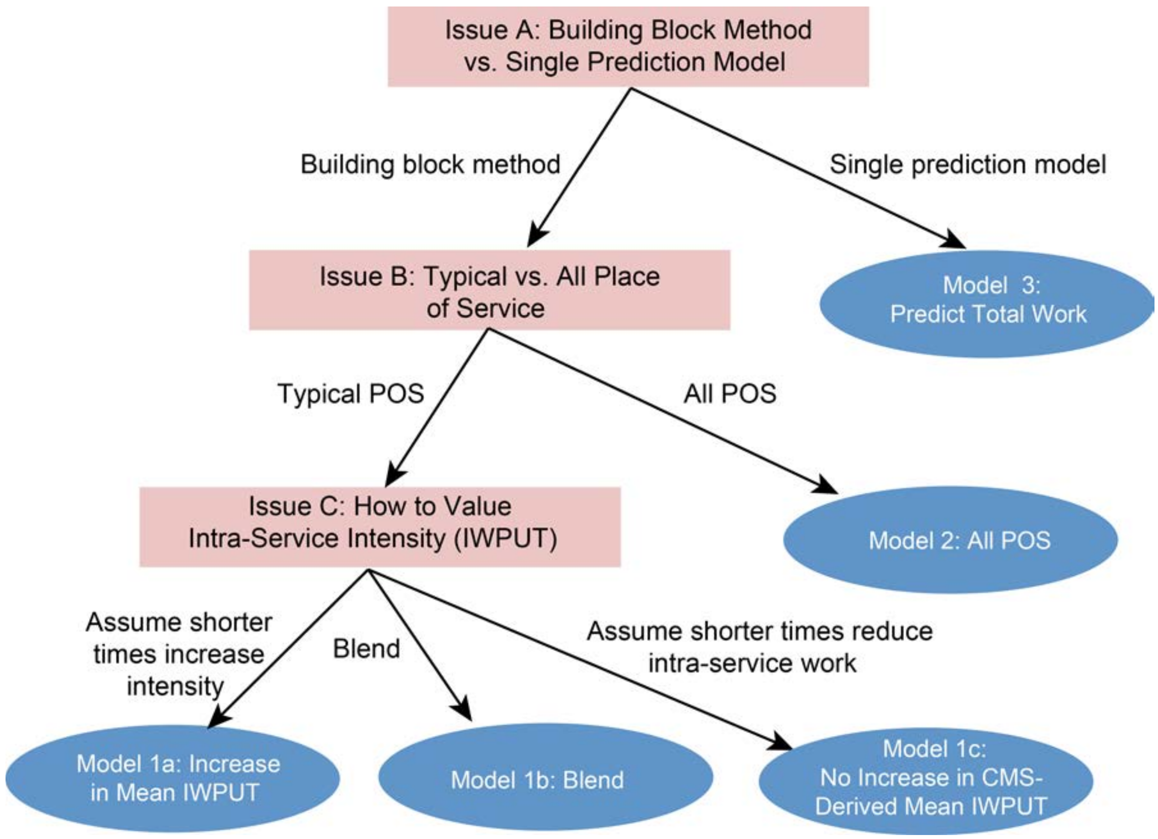
There are two arguments for increasing mean IWPUT for surgical procedures. The first is that the current RUC process for valuation focuses on total work. The valuation of subcomponents does not receive as much scrutiny. Intra-service work and IWPUT are derived values, and inaccuracies in other subcomponents (in particular, post-operative E&M visits) make the intra-service work values suspect. It is possible that IWPUT is currently underestimated in the RUC process and therefore in the CMS estimates. Another argument for increasing IWPUT is that while technology and efficiency gains mean surgical procedures are getting shorter in duration, work per unit time is correspondingly higher. For example, while a colonoscopy can be done more quickly, the argument is that work per unit time during colonoscopy in higher because physicians must maintain a higher level of mental focus and use higher levels of technical skill.

The decision on mean IWPUT has critical implications for valuing intra-service work and therefore total work. However, the lack of a gold standard for intensity means that RAND cannot empirically assess which argument has more basis. We therefore model both.[29]

---

[29] The RAND time estimates are not uniformly lower than the CMS time estimates across all procedures. For many procedures, the RAND times are the same or longer so that creating a uniform adjustment to IWPUT across all procedures may be problematic. Therefore, we create a procedure-specific adjustment for IWPUT.

Figure 7.4. Overview of RAND Models



Our three models illustrate a range of arguments. In Model 1a, IWPUT is calculated at the very end of the analysis by dividing predicted intra-service work by RAND intra-service times. In Model 1c, we predict IWPUT directly using the CMS time file values ($IWPUT_{CMS\ corrected}$) as the dependent variable. In Model 1b, we blend the two approaches. There are numerous approaches for building models for intra-service work and IWPUT. The Model 1 options illustrate one strategy, but other approaches may be employed that have the same general effect.

Model 1a: Increased IWPUT

In Model 1a, we use regression models to predict intra-service work using procedure characteristics described above:

$$\text{Derived intra-service work}_{ij} = \beta_0 + \beta_1 \ln(time_{RAND,typical})_{ij} + \mathbf{X}_i\boldsymbol{\gamma} + \mathbf{Z}_{i,\text{typical}}\boldsymbol{\delta} + \mathbf{W}_{i,\text{typical}}\boldsymbol{\psi} + u_j + \varepsilon$$

We model derived intra-service work as a function of log-transformed RAND intra-service time ($\ln(time_{RAND})$), procedure characteristics (X_i), measures of patient complexity (Z_i), measures of procedure complexity and risk (W_i), a code group random effect u_j, and an error term.[30] Regression coefficients for this model and the other models are in Appendix G.[31]

Because this regression model is based on the derived values, these prediction models are, in essence, recalibrating or changing the values across the codes. The average predicted intra-service work across the codes will be similar. We use the predicted values from these regressions paired with RAND times to "back out" IWPUT. Because RAND time estimates are shorter on average, mean IWPUT increases.

Model 1c: No Change in Mean IWPUT Value

Under Model 1c, we shift from predicting intra-service work to predicting IWPUT based on the CMS estimates. Predicting IWPUT might be seen as more consistent with the BBM approach, where each component is independently estimated and built up to estimate total work.

$$\text{CMS IWPUT}_{ij} = \beta_0 + \beta_1 \ln(time_{RAND,typical})_{ij} + \mathbf{X}_i\boldsymbol{\gamma} + \mathbf{Z}_{i,typical}\boldsymbol{\delta} + \mathbf{W}_{i,typical}\boldsymbol{\psi} + u_j + \varepsilon$$

We model the CMS IWPUT with a floor of 0.0224 ($IWPUT_{ij}$) as a function of log-transformed RAND intra-service time ($\ln(time_{RAND})$), procedure characteristics (X_i), measures of patient complexity (Z_i), measures of procedure complexity and risk (W_i), a code group random effect u_j, and an error term.[32]

Model 1b: Blend

In Model 1b, we blend the Model 1a and Model 1c IWPUT and intra-service work values by averaging across the 1a and 1c IWPUT values and then multiplying the result by RAND time.

A simplified example might be the easiest way to illustrate this choice. Say the Model 1a IWPUT is 0.06, the Model 1c IWPUT is 0.02, and the RAND intra-service time is 20 minutes. The Model 1b IWPUT is the average of these IWPUT values (0.04), and the Model 1b intra-service work is the product of 0.04 and the RAND time estimate of 20 minutes, or 0.8 RVUs.

Model 2: All POS

Model 2 takes a very similar approach to Model 1a. The only difference is that the intra-service work prediction model now focuses on all POS, not just the typical POS.

---

[30] See Appendix C for a description of the specific variables included in X, Z, and W). We estimate the coefficients for the model as gamma GLM with a log link using Stata 13's meglm command.

[31] The intra-service work model using procedure characteristics defined at the typical place of service (Model 1a) accounts for 71.9 percent of the variation in the underlying variation in our adjusted intra-service work estimate. See Chapter 5 for details on our approach to calculating fraction of variance explained and root mean square error.

[32] See Appendix C for a description of the specific variables included in X, Z, and W). We estimate the coefficients for both models as mixed-effect gamma GLM with a log-link (using Stata 13's meglm command).

$$\text{Intra-service work}_{ij} = \beta_0 + \beta_1 \ln(time_{RAND,all})_{ij} + \mathbf{X}_i\boldsymbol{\gamma} + \mathbf{Z}_{i,all}\boldsymbol{\delta} + \mathbf{W}_{i,all}\boldsymbol{\psi} + u_j + \varepsilon$$

Results of Prediction Models

Model 3 focuses on total work and does not generate a separate estimate for IWPUT and intra-service work. We therefore only present the findings in this section for Model 1and Model 2.

IWPUT

The mean predicted IWPUT from each model is presented in Table 7.4. As discussed above, the mean predicted IWPUT in Model 1a and Model 2 is 37 and 32 percent, respectively, higher than in the CMS time file (across codes, not weighted by service volume). Model 2 results are similar to Model 1a. The choice of all POS versus typical POS has little impact at an aggregate level. Mean predicted IWPUT in Model 1c is only 3 percent higher than the mean CMS time file IWPUT. This small increase is due to our use of the 0.0224 RVU/minute floor. Mean predicted IWPUT in Model 1b is halfway between Models 1a and 1c (a 19-percent increase over the time file mean).

Table 7.4. Mean Predicted IWPUT and Intra-Service Work Across Models

	Mean IWPUT (RVUs/minute)	Percentage of IWPUT Values That Are Outside of Range^	Mean Intra-Service Work (RVUs)
CMS time file	0.057	15.5	4.8
Model 1a	0.078	9.1	5.3
Model 1b	0.068	2.1	4.8
Model 1c	0.059	0.6	4.2
Model 2	0.075	6.7	5.1

^ As noted above, one concern with the CMS IWPUTs is that they lack face validity. To create an IWPUT range, we took the CMS IWPUT for most pre-service and immediate post-service work, which is 0.0224, and the IWPUT in the current time file for a very complex procedure, CPT 32854 (lung transplant with bypass), which is 0.11.

In contrast to the CMS time file, across all our models a very small fraction of procedures have an IWPUT that is outside the IWPUT range we created (Table 7.4). There are no negative IWPUT values across the RAND models.

There is a differential impact based on volume of procedures. Among low-volume procedures there is an increase in IWPUT, but among high-volume codes there is a drop in IWPUT. In Table 7.5, we show IWPUT across models *weighted* by volume. In the weighted results, mean IWPUT and intra-service work are lower than in the unweighted results. This is particularly true for Model 1c, where the overall weighted mean predicted IWPUT is 22.4 percent lower than time file values; however, the weighted mean increases 7.1 percent for the procedures performed less than 1,000 times per year, compared to a 28.2 percent reduction for procedures performed more than 100,000 times per year. This differential change in IWPUT across the volume categories is driven by differences in intra-service time between RAND and CMS estimates. Across the procedures, RAND intra-service time is substantially lower *except* among high-volume codes.

There are also some unusual results. For example, in Model 1a the weighted mean IWPUT value for the procedures that focus on the ear increases 39 percent, while the value for procedures that focus on the eye decreases 23.4 percent. We discuss in Chapter 10 the clinical input we obtained on IWPUT and some complexities in how it can be modeled. In Table 7.6, we show the different estimates of IWPUT across 20 of the most common procedures.

Table 7.5. Percentage Difference Between CMS Volume-Weighted Mean IWPUT Estimates and RAND Estimates, by Procedure Category

	CMS Estimate Mean RVUs	Model 1a % Difference	Model 1b % Difference	Model 1c % Difference	Model 2 % Difference
Total	0.09	1.31	–10.54	–22.38	–0.92
CMS intra-service time categories					
0 to 30 minutes	0.09	–9.96	–18.44	–26.93	–11.97
31 to 70 minutes	0.07	33.94	11.77	–10.39	32.07
71 to 120 minutes	0.07	20.27	3.46	–13.34	15.01
Over 120 minutes	0.07	21.26	7.82	–5.61	15.63
Global period					
0	0.09	–5.87	–15.15	–24.42	–7.07
10	0.05	17.47	6.04	–5.40	15.75
90	0.10	19.40	–1.23	–21.87	13.68
Not applicable	0.04	63.54	38.93	14.33	53.70
Typical place of service					
ED	0.05	50.82	38.31	25.80	41.39
Inpatient	0.09	–9.91	–16.54	–23.16	–9.56
Office	0.09	–13.95	–20.81	–27.68	–16.05
Ambulatory facility (HOPD or ASC)	0.09	25.94	4.74	–16.46	22.59
Risk category					
Office-based	0.09	–16.82	–21.89	–26.95	–18.20
ASC	0.09	16.81	–1.59	–19.99	13.93
Hospital outpatient	0.09	19.60	6.67	–6.26	18.51
Inpatient only	0.08	7.12	–3.88	–14.88	0.71

	CMS Estimate Mean RVUs	Model 1a % Difference	Model 1b % Difference	Model 1c % Difference	Model 2 % Difference
Body system grouping					
Nervous system	0.07	12.27	3.65	–4.97	9.64
Endocrine system	0.06	25.69	5.47	–14.76	13.67
Eye	0.18	–23.39	–32.48	–41.57	–25.15
Ear	0.05	38.93	29.00	19.07	32.66
Nose, mouth, and pharynx	0.09	–21.12	–28.03	–34.94	–25.11
Respiratory system	0.08	–7.26	–12.30	–17.35	–11.66
Cardiovascular system	0.09	3.75	–3.63	–11.01	3.58
Hemic and lymphatic system	0.06	32.29	21.45	10.60	33.19
Digestive system	0.08	29.72	4.64	–20.44	28.24
Urinary system	0.08	12.02	–1.94	–15.90	5.03
Male genital organs	0.08	13.09	–2.36	–17.82	10.27
Female genital organs	0.06	27.61	14.75	1.89	21.09
Musculoskeletal system	0.09	–7.90	–16.00	–24.11	–10.03
Integumentary system	0.04	52.25	31.81	11.38	49.88
Miscellaneous services	0.08	30.93	14.58	–1.78	28.93
Number of annual Medicare procedures					
Less than 1,000	0.05	43.91	25.52	7.14	38.28
1,000 to 9,999	0.06	33.70	17.87	2.04	29.46
10,000 to 99,999	0.07	23.07	8.06	–6.95	19.43
100,000 or more	0.10	–6.73	–17.45	–28.17	–8.44

Table 7.6. Percentage Difference Between Derived CMS IWPUT Values and RAND Estimates, "Top 20" Codes

Code	Description	CMS IWPUT Estimate	Model 1a % Difference	Model 1b % Difference	Model 1c % Difference	Model 2 % Difference
13132	Complex laceration repair	0.07	–10.4	–20.5	–30.6	–10.7
17311	Mohs procedure of skin	0.05	22.9	14.3	5.6	25.5
20610	Drain or inject a bursa or joint	0.10	–20.7	–25.7	–30.7	–21.4
27245	Treat thigh fracture	0.09	4.7	–13.2	–31.2	–6.1
27447	Total knee arthroplasty	0.11	–11.3	–22.8	–34.3	–18.0
31231	Diagnostic nasal endoscopy	0.12	–35.6	–39.4	–43.1	–38.1
33533	CABG arterial single	0.10	–8.9	–4.4	0.1	–8.7

Code	Description	CMS IWPUT Estimate	Model 1a % Difference	Model 1b % Difference	Model 1c % Difference	Model 2 % Difference
35301	Rechanneling of artery	0.12	–10.7	–18.6	–26.4	–17.6
43239	Esophagogastroduodenoscopy (EGD) with biopsy	0.11	–15.2	–26.8	–38.4	–14.9
44120	Removal of small intestine	0.04	113.6	87.0	60.5	117.4
45380	Colonoscopy with biopsy	0.06	142.5	80.2	17.8	139.8
47562	Laparoscopic cholecystectomy	0.08	17.1	–0.5	–18.0	7.0
52000	Cystoscopy	0.11	–5.0	–18.0	–31.0	–11.3
52601	Prostatectomy (TURP)	0.09	32.8	4.7	–23.4	22.1
62311	Inject spine lumbar/sacral	0.06	39.2	28.0	16.8	33.7
63047	Remove spinal lamina (lumbar)	0.08	7.8	–2.6	–12.9	1.6
64450	Digital nerve block	0.08	–13.0	–17.1	–21.2	–12.3
66984	Cataract surgery with intraocular lens (1 stage)	0.21	–7.7	–26.3	–44.9	–12.3
67228	Treatment of retinal lesion	0.15	–1.0	–17.7	–34.5	–0.1
93458	Left heart artery and ventricle angiography	0.09	15.3	8.5	1.7	26.5

Intra-Service Work

In Figure 7.5, we show a scatter plot comparing CMS intra-service work to predicted intra-service work from our Model 1a. The most notable change is that there are no negative intra-service work estimates in the RAND model.

Similar to our findings in IWPUT, our unweighted and weighted results show different patterns. In Tables 7.7, we show the weighted mean change in predicted intra-service work across the RAND models, while the unweighted results are in Appendix H. In Table 7.7, the Model 1a and 2 values are on average slightly lower than the CMS time file values (–2.9 percent and –4.8 percent, respectively). Defining covariates in the typical POS or across all POS (Model 1a versus Model 2) appears to make little difference in aggregate.

In Model 1c, intra-service work is 25 percent lower on average than in CMS time. In Model 1b, intra-service work values are halfway between the Model 1a and 1c results.

Across categories of procedures, there are considerable changes in intra-service work. Surgical procedures associated with a 90-day global period have 5.6- to 32.0-percent lower predicted values on average compared to CMS values. Across the body system groupings, there are both substantial increases (e.g., ear and integumentary systems) and decreases (eye and nose, mouth, and pharynx systems). Across the categories of volume of care, the highest-volume procedures have an average reduction in intra-service work while the lowest-volume procedures have the highest average increases. In Table 7.8, we show the different estimates of intra-service work across 20 of the most common procedures.

Figure 7.5. Comparison of CMS and Predicted Model 1a Intra-Service Work

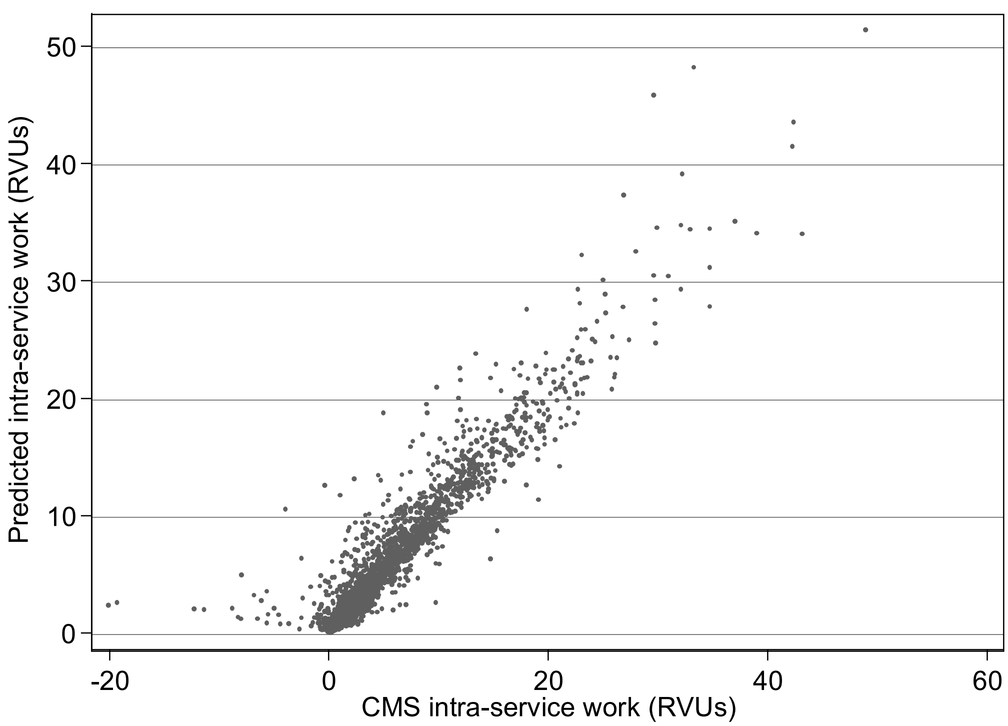


Table 7.7. Percentage Difference Between CMS Volume-Weighted Mean Intra-Service Work and RAND Estimates, by Procedure Category

	CMS Estimate Mean RVUs	Model 1a % Difference	Model 1b % Difference	Model 1c % Difference	Model 2 % Difference
Total	2.21	–2.93	–13.81	–24.69	–4.79
CMS intra-service time categories					
0 to 30 minutes	1.01	13.47	3.24	–6.99	12.64
31 to 70 minutes	3.17	–11.83	–24.09	–36.36	–12.89
71 to 120 minutes	7.24	–10.26	–21.51	–32.77	–13.59
Over 120 minutes	13.08	–5.31	–13.32	–21.34	–9.16
Global period					
0	1.52	0.45	–8.53	–17.51	–0.38
10	1.35	–9.85	–18.28	–26.72	–8.48
90	5.54	–5.56	–18.78	–32.00	–8.94
Not applicable	0.99	–1.36	–11.41	–21.47	–3.39
Typical place of service					
ED	1.04	127.19	110.80	94.40	121.46
Inpatient	4.71	–3.72	–11.66	–19.60	–8.32
Office	1.08	9.48	0.52	–8.44	9.60
Ambulatory facility (HOPD or ASC)	2.79	–12.15	–26.09	–40.04	–12.97
Risk category					
Office-based	0.93	22.47	16.17	9.87	22.47
ASC	2.62	–12.42	–25.81	–39.19	–12.98
Hospital outpatient	4.53	–2.50	–11.43	–20.35	–9.43
Inpatient only	10.97	–0.84	–9.66	–18.49	–6.54
Body system grouping					
Nervous system	1.28	10.16	0.29	–9.58	3.99
Endocrine system	7.16	–8.35	–23.57	–38.79	–14.13
Eye	2.69	–0.11	–12.26	–24.40	–0.12
Ear	1.82	11.32	1.46	–8.39	9.47
Nose, mouth, and pharynx	1.45	–14.98	–24.18	–33.38	–17.18
Respiratory system	1.65	–5.43	–10.14	–14.84	–10.29
Cardiovascular system	4.18	–3.80	–9.44	–15.08	–6.03
Hemic and lymphatic system	4.33	–0.17	–7.73	–15.30	–0.28
Digestive system	3.09	–23.59	–36.53	–49.48	–24.25

	CMS Estimate Mean RVUs	Model 1a % Difference	Model 1b % Difference	Model 1c % Difference	Model 2 % Difference
Urinary system	2.13	–9.18	–19.94	–30.71	–10.67
Male genital organs	3.94	–5.04	–19.52	–34.00	–9.84
Female genital organs	2.85	11.25	–1.33	–13.92	3.85
Musculoskeletal system	1.70	11.83	–0.89	–13.62	6.87
Integumentary system	1.71	7.37	–6.23	–19.83	9.69
Miscellaneous services	0.78	–12.17	–22.51	–32.85	–13.18
Number of annual Medicare procedures					
Less than 1,000	5.44	13.60	1.83	–9.95	9.62
1,000 to 9,999	4.38	2.89	–7.82	–18.52	–0.57
10,000 to 99,999	3.06	–5.65	–16.44	–27.24	–7.39
100,000 or more	1.59	–3.73	–14.65	–25.58	–5.06

Table 7.8. Percentage Difference between CMS Intra-Service Volume-Weighted Mean Estimates and RAND Estimates, "Top 20" Codes

Code	Description	CMS Estimate Mean RVUs	Model 1a % Difference	Model 1b % Difference	Model 1c % Difference	Model 2 % Difference
13132	Complex laceration repair	3.7	–41.0	–47.7	–54.3	–40.5
17311	Mohs procedure of skin	5.6	–8.7	–15.2	–21.6	–7.3
20610	Drain or inject a bursa or joint	0.5	66.9	56.4	45.9	67.0
27245	Treat thigh fracture	6.8	–10.8	–26.1	–41.4	–20.1
27447	Total knee arthroplasty	11.0	–12.4	–23.7	–35.0	–18.9
31231	Diagnostic nasal endoscopy	0.9	–40.5	–43.9	–47.4	–41.6
33533	CABG arterial single	15.1	11.6	17.1	22.7	11.7
35301	Rechanneling of artery	14.5	–17.4	–24.7	–31.9	–23.8
43239	Esophagogastroduodenoscopy (EGD) with biopsy	1.7	–39.8	–48.0	–56.2	–42.5
44120	Removal of small intestine	5.1	52.5	33.6	14.6	55.1
45380	Colonoscopy with biopsy	2.9	–20.6	–41.0	–61.4	–22.0
47562	Laparoscopic cholecystectomy	6.6	–24.9	–36.2	–47.4	–23.8
52000	Cystoscopy	1.7	–35.1	–43.9	–52.8	–34.3
52601	Prostatectomy (TURP)	6.6	–8.3	–27.7	–47.1	–12.6
62311	Inject spine lumbar/sacral	0.6	43.9	32.4	20.8	33.4
63047	Remove spinal lamina (lumbar)	7.0	12.7	1.9	–8.9	4.2
64450	Nerve block, other peripheral	0.4	85.4	76.8	68.1	67.4
66984	Cataract surgery with intraocular lens (1 stage)	4.4	–23.5	–38.9	–54.3	–24.1
67228	Treatment of retinal lesion	9.1	–51.0	–59.3	–67.6	–50.9
93458	Left heart artery and ventricle angiography	4.2	–9.9	–15.3	–20.6	–22.9

Summary and Implications for Our Model

In this chapter we explored various options on how to model intra-service work and IWPUT. As noted in Chapter 4, our RAND time estimates are lower on average than the CMS estimates.

One key issue explored in this chapter is what this implies for IWPUT and intra-service work. In some models, we assume that intra-service work remains constant, and therefore predicted IWPUT is higher on average. In another model, we assume that IWPUT stays the same on average, and therefore intra-service work decreases. In a middle scenario we blended changes in IWPUT and intra-service work. This has important implications for estimated total work RVUs as will be illustrated in the following chapters.

Limitations

One key limitation RAND faced in modeling intra-service work and IWPUT is the lack of a gold standard on intensity. Because of this, we used derived intra-service work and IWPUT values that use the reverse BBM to develop our models. Because these models are based the derived values, the RAND models are in essence recalibrating or changing the values across the codes. Ideally, the prediction models would be based on an independent measure for intensity or some other mechanism to estimate intensity. Another argument for using a different strategy to estimate intensity is that current time file IWPUT and our derived IWPUT values have little correlation with variables that at least on face value should be associated with IWPUT. Much of the current variation in IWPUT in the CMS time file is unexplained and might represent noise.

One alternative strategy for estimating IWPUT is to create intensity "buckets" in which there would be some prespecified number of intensity categories and procedures would be assigned to those buckets based on their clinical characteristics. Such a strategy would be consistent with the other aspects of the building block approach (pre-service work, immediate post-service work, post-operative E&M visits), which all have been given a single constant intensity. It would also be consistent with the current RUC approach for pre-service work, which assigns procedures to a set of pre-specified pre-service "bundles." Such an exercise would require significant clinical input, as right now there is no clear method to assign the IWPUT.

Another option we have considered is creating independent estimates of intensity. In the future directions section of Chapter 10, we describe future work in which we hope to consider other means of estimating intensity. Having physicians estimate intra-service work directly is one strategy to help generate potentially more accurate estimates of IWPUT than are in the current time file and is more consistent with the approach taken in the original Harvard model.

Another key limitation is that we might be using the wrong characteristics to predict intra-service work and IWPUT. The characteristics we use are consistent with the prior theoretical work done by the Harvard team in developing the RBRVS. More recent work by Horner and Jacobsen (2011a, 2011b) in assessing intensity has found that intensity is also related to such issues as challenging patients, schedule pacing, uncertainty in decisionmaking, and physician strain. These variables are hard to capture using external databases and may not even be procedure-specific. For example, schedule pacing will vary from day to day and is more likely to be practice-specific than procedure-specific.

We made many choices in creating our predictive models. For example, to address nonsensical intra-service work values where the value is negative, we created a minimum IWPUT value before creating our predictive models. There are other choices we could make in creating the models, which may result in different findings.

Implications for Model

The IWPUT and intra-service work results are key inputs on our models of total work. Which approach to take depends on theoretical and policy considerations. We have outlined the arguments for the different model approaches. The choice of the model is a critical one, as it can have significant impact on total work estimates. These issues are further explored in Chapter 8.

8. Estimating Total Work RVUs

Overview

In prior chapters, we explored various options for estimating the individual work components using the BBM. In this chapter, we focus on total work RVUs. For Models 1 and 2, we combine the results from the earlier chapters to estimate total work RVUs using the BBM. In Model 3, we use a single prediction model to predict total work RVUs using the same type of variables as we did for intra-service work. We start with an overview of the models and how they vary. We then provide the results and summary statistics that can be used to compare the models. We conclude by discussing their relative strengths and weaknesses.

Summary of Model Options

In prior chapters, we investigated different approaches and assumptions to modeling the individual components of total work RVUs using the BBM. Table 8.1 summarizes how we have chosen to combine these modeling options into models for estimating total work in Models 1 and 2. Model 3 does not use the BBM but rather uses a single prediction model to estimate total work RVUs.

Table 8.1. Summary of Modeling Choices Reflected in Report Models

Issue	Model 1a	Model 1b	Model 1c	Model 2	Model 3
Issue A—Method	BBM	BBM	BBM	BBM	Single prediction model for total work
Issue B—Setting	Typical	Typical	Typical	All POS	Typical
Issue C—How to value IWPUT	$IWPUT_{RAND-P}$ = Predicted intra-service work ÷ RAND time	$IWPUT_{BLEND}$ = $IWPUT_{RAND-P}$ x 0.5 + $IWPUT_{CMS-P}$ T x 0.5	$IWPUT_{CMS-P}$ = predicted values based on values derived from CMS estimates	$IWPUT_{RAND-P}$ = Predicted intra-service work ÷ RAND time	$IWPUT_{BLEND}$ = RAND IWPUT x 0.5 + CMS IWPUT x 0.5
How are total work RVUs calculated?	Pre-service $work_{predicted}$ + intra-service $work_{predicted}$ + post-service $work_{predicted}$ + post-operative $E\&M_{predicted}$	Pre-service $work_{predicted}$ + $IWPUT_{BLEND}$ x RAND time + post-service $work_{predicted}$ + post-operative $E\&M_{predicted}$	Pre-service $work_{predicted}$ + $IWPUT_{CMS-P}$ x RAND time + post-service $work_{predicted}$ + post-operative $E\&M_{predicted}$	Pre-service $work_{predicted}$ + intra-service $work_{predicted}$ + post-service $work_{predicted}$ + post-operative $E\&M_{predicted}$	Predicted value

In Chapter 3, we discussed whether our estimates should be based on the typical setting or the full range of settings in which services are provided. This decision has a direct effect on our estimates for intra-service time, pre-service, immediate post-service, and post-operative visits. It also has an indirect effect on the derived values for intra-service work. Model 2 aggregates the RAND estimate for each work component based on the full range of settings in which services are provided (all POS). Models 1 and 3 are based on the typical POS.

In Chapter 7, we discussed options for estimating intra-service work using the BBM. For Model 1, we investigate alternative ways to derive the initial intra-service work values. In Model 1a, we derive intra-service work by subtracting the predicted values for the other work components from total work RVUs and assume that the changes in intra-service times do not affect intra-service work values. In Model 1b, we derive an IWPUT value that assumes that half of the time difference affects intensity and that the other half affects intra-service work. In Model 1c, we assume that the changes in intra-service time affect intra-service work but not IWPUT. We predict an IWPUT value based on values derived from CMS estimates. We estimate intra-service work in Models 1b and 1c by multiplying the respective IWPUT values by RAND intra-service times.

Methods for Calculating Predicted Total Work

For Models 1 and 2, we calculate predicted total work by adding together our predicted values for each BBM component. The specific components include:

- predicted pre-service work (the sum of predicted values for evaluation, pre-service positioning, and scrub work)
- predicted intra-service work
- predicted immediate post-service work
- predicted post-operative E&M visit work (the sum of predicted values for inpatient visits work, outpatient visit work, critical care visit work, and discharge visit work).

We predict total work directly in Model 3. Our total work prediction model is very similar to the intra-service work prediction models discussed in Chapter 7. The dependent variable is the total work RVUs in the CMS estimates. The independent variables include log-transformed RAND intra-service time, procedure characteristics, measures of patient complexity, measures of procedure complexity and risk, a code group random effect, and an error term.[33] The regression results are reported in Appendix G. As in previous chapters, we generate predicted values using estimated coefficients and each procedure's values for the independent variables.

Table 8.2 summarizes the mean total work RVUs from the CMS estimates and from our prediction model. There are a large number of high-volume procedures with relatively low work

[33] See Appendix C for a description of the specific variables included in the regression. We estimate these models as gamma GLM with a log link using Stata 13's meglm command.

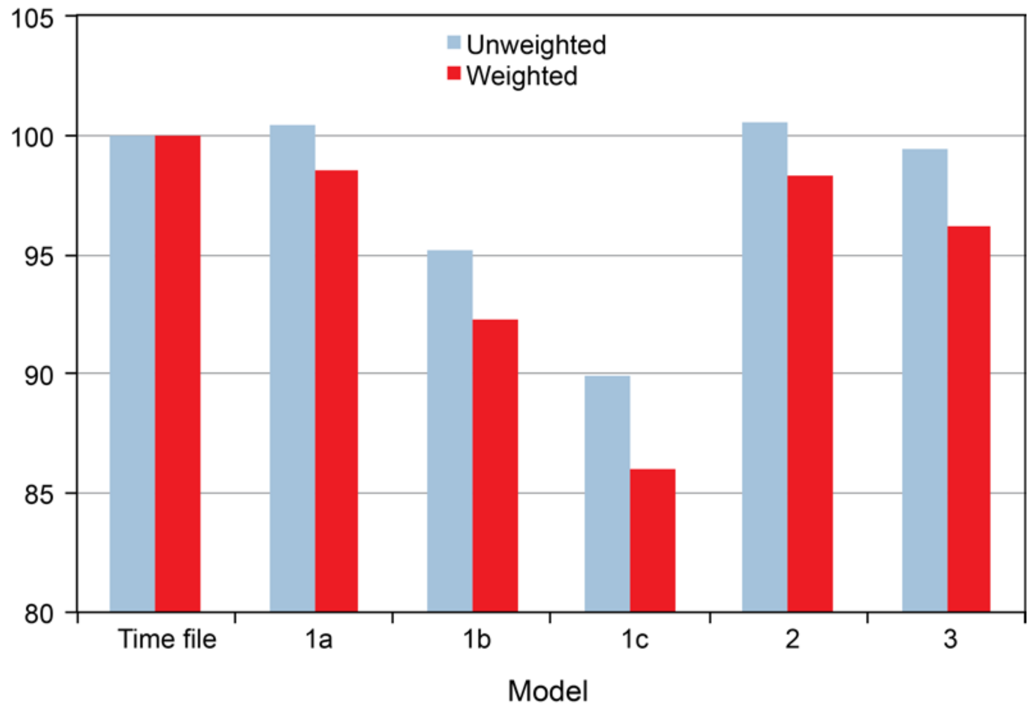
RVUs, so that the volume-weighted means for both the CMS estimates and predicted values are much lower on average than the unweighted means. The Model 3 results are discussed further in the next section.

Table 8.2. Comparison of CMS Total Work RVUs and RAND Model 3 Estimates

	CMS Estimates	Predicted Value
Unweighted mean	10.54	10.48
% difference from CMS	-	–0.5%
Mean weighted by 2012 Medicare volume	3.84	3.69
% difference from CMS	-	–3.9%

Figure 8.1 shows the mean total work RVUs under each of the models relative to the CMS values. The Model 1a and 2 predicted values were 1 percent less on average compared to CMS values. The difference in weighted means is slightly greater: 1.4 percent lower for Model 1a and 1.7 percent lower for Model 2. The unweighted mean RVUs for Model 3 are 0.5 percent lower than the CMS value, but the weighted means are 3.8 percent lower. Reflecting the reductions made in intra-service work for shorter intra-service times, the unweighted mean RVUs for Models 1b and Model 1c are 4.8 percent and 10.0 percent lower, respectively, than the CMS unweighted mean. The differences in the weighted means are larger: Model 1b is 7.8 percent lower, and Model 1C is 14.0 percent lower.

Figure 8.1. Mean Total Work RVUs Predicted by Models Relative to CMS Values

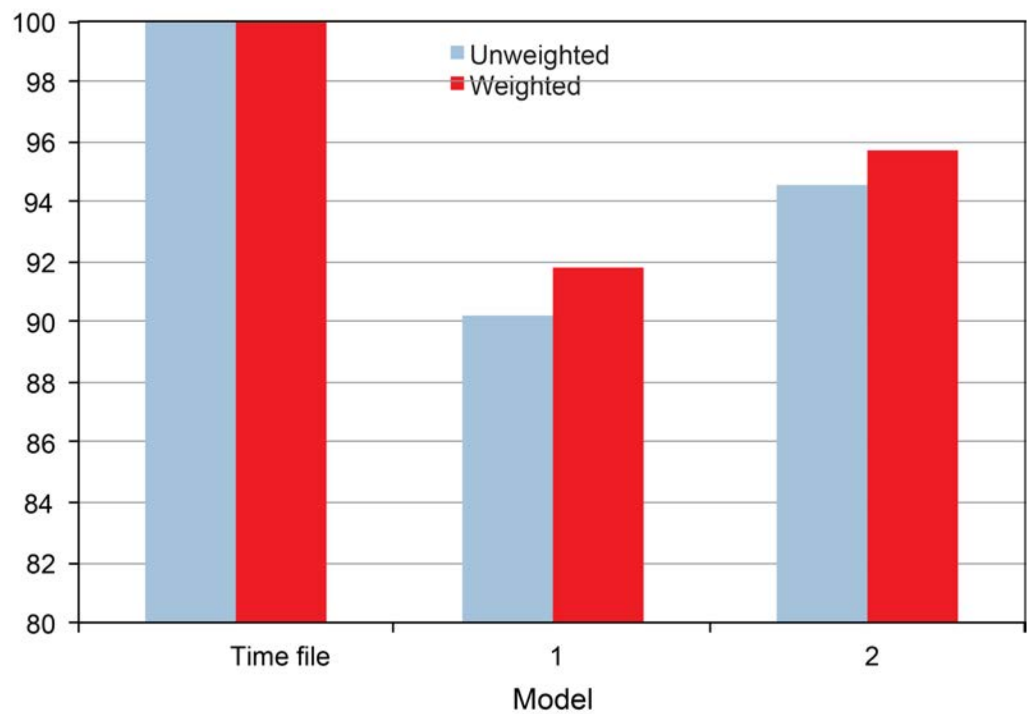


The overall differences in predicted total work across the models are due to changes in several of the work components: constraining statistical outliers in predicting pre-and immediate post-service work, correcting the E&M visit counts to conform to the typical setting and the median length of stay between the date of the procedure and discharge for inpatient procedures, and adjusting intra-service work values to account for the shorter intra-service times. The higher unweighted means in Models 1a and 2 (0.5 percent and 0.6 percent, respectively) reflect the floor placed on minimum intra-service work values. The higher Model 1b and 1c mean values reflect adjustments to IWPUT values for shorter intra-service times. As discussed in Chapter 7, these results are expected. The Model 1c assumes the changes in intra-service time affect intra-service work but not intensity. The intra-service work estimates are calculated by multiplying the RAND times by IWPUT predicted from CMS values that essentially hold the mean IWPUT value constant. Model 1b is a blend of Model 1a, which essentially assumes no changes in intra-service work, and Model 1c.

In the aggregate, the pre-service and immediate post-operative RVUs are similar under each of the BBM models (data not shown). There is only a slight reduction in the unweighted mean RVUs for these work components across the four models. In contrast, the weighted means increase 11–12 percent, largely because of the large percentage increases in high-volume codes with low work values. When the CMS RVUs are relatively low, a relatively minor change in the absolute RVU values results in a large percentage change. For example, the CMS weighted mean work RVUs for the pre-service and immediate post-service components for services typically furnished in the emergency department is 39 percent higher than the CMS weighted mean RVUs (Table 5.9), but the absolute difference is only 0.15 RVUs.

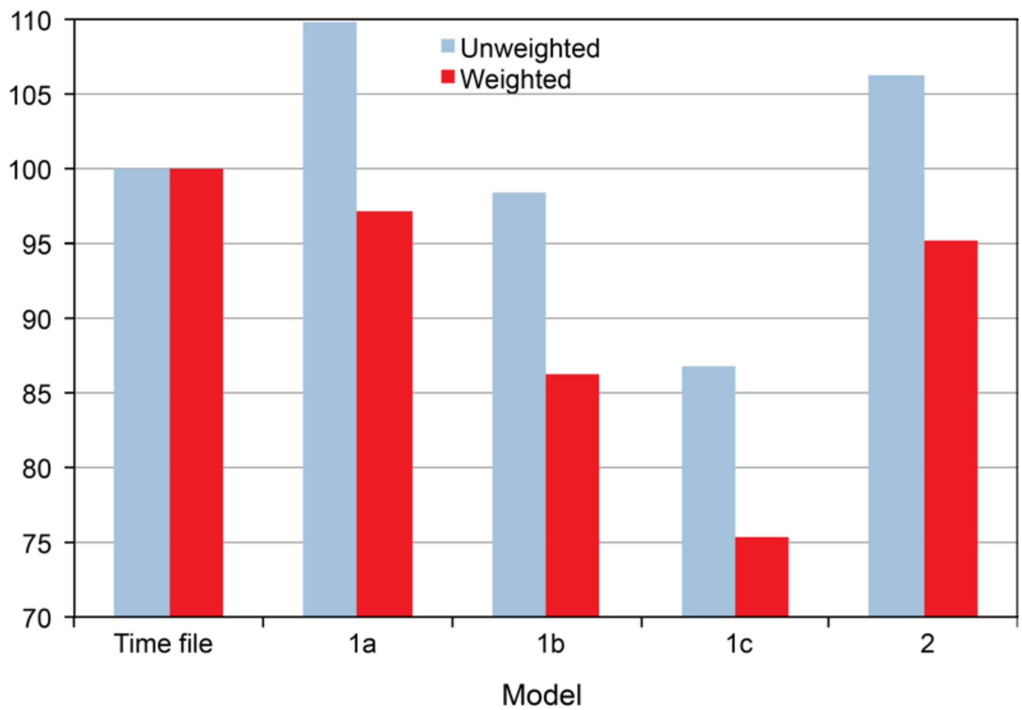
There are two sets of corrections in the prediction models for post-operative visits: those for typical POS (Model 1) and those for all POS (Model 2). The difference between CMS post-operative work and predicted post-operative work is larger in Model 1 than in Model 2 (Figure 8.2). The unweighted mean difference for Model 1 is 10.8 percent versus 5.5 percent for Model 2. The weighted mean difference is 8.2 percent for Model 1 versus 4.3 percent for Model 2. This pattern is consistent with how the all POS corrections were made. Relative to the typical POS, the E&M visits counts for all POS are lower when the typical setting is inpatient (because the inpatient visits are dropped for the proportion of procedures performed in ambulatory settings) and higher when the typical setting is ambulatory (because an estimate of E&M inpatient visits is added for the proportion of procedures performed in the inpatient setting). A large proportion of the procedures performed in ambulatory settings are high-volume procedures. The smaller reduction in the weighted means than in the unweighted means seen in both Models 1 and 2 is also consistent with the corrections being made primarily to procedure codes with 10- or 90-day global periods and relatively high work RVUs.

Figure 8.2. Mean RVUs for Post-Operative Visits Predicted by Models Relative to CMS Values



There are larger differences across the predicted values for intra-service work than for the other work components (Figure 8.3) and they drive the changes in total work RVUs seen earlier in Figure 8.1. The increases in the Model 1a and Model 2 estimates are attributable to three factors: (1) the lower post-operative visit work is offset by increases in derived intra-service work values, (2) the floor on derived intra-service work values used in the prediction models, and (3) no adjustment to intra-service work for the changes in intra-service times. The Model 1b and 1c are lower than the CMS estimates because intra-service work is adjusted for the half of the intra-service times change in Model 1b and for all of the change in Model 1c. Relative to the CMS estimates, the reductions are greater in the Model 1b and 1c weighted means than in the unweighted means. This is consistent with the pattern of increased IWPUT values for lower-volume procedures and reduced IWPUT values for higher-volume procedures seen in Table 7.5.

Figure 8.3. Mean Intra-Service Work RVUs Predicted by Models Relative to CMS Values



In Table 8.3, we show the percentage differences in the total work RVUs estimated under each model relative to current CMS estimates by procedure characteristics. The differences shown in this table are weighted by Medicare volume; unweighted results are reported in Appendix H. A comparison of Model 1a to Model 3 provides a sense of whether the prediction models for each work component using the BBM approach produces comparable results to predicting total work in a single prediction model. Using our methodologies, they produce somewhat different results. The overall percentage difference for Model 1a is –1.44 percent, compared to –3.9 percent for Model 3. With few exceptions (mostly in the body system categories—e.g., digestive, urinary, and male genital organ systems), the Model 3 mean total RVUs across procedure categories are also consistently lower than the Model 1 means. However, the magnitude of the Model 3 changes compared to the Model 1differs across procedure categories. Using the typical place of service categories as an example, the reduction in the weight mean RVUs in Model 1a and Model 3 for ambulatory facility services are similar (–9.1 percent versus –9.5 percent), but there is a marked difference for ED procedures (92.8 percent versus 49.3 percent). Separate predictions for each work component may allow for a greater range of total values than the single prediction model (see Figure 8.4). Compared to Model 1a, the distribution of procedure codes under Model 3 is smoother and somewhat more concentrated in the lower work RVUs.

Table 8.3. Percentage Difference in Total Work RVUs Relative to CMS Total Work RVUs, Weighted by Procedure Volume

	CMS Estimate Mean RVUs	Model 1a % Difference	Model 1b % Difference	Model 1c % Difference	Model 2 % Difference	Model 3 % Difference
Total	3.84	–1.44	–7.72	–13.99	–1.70	–3.85
CMS intra-service time categories						
0 to 30 minutes	1.76	12.02	6.18	0.34	11.58	6.80
31 to 70 minutes	5.14	–8.41	–15.97	–23.54	–9.68	–9.75
71 to 120 minutes	12.85	–7.92	–14.26	–20.60	–6.79	–8.89
Over 120 minutes	25.06	–4.37	–8.56	–12.74	–4.16	–5.46
Global period						
0	2.12	3.99	–2.45	–8.88	3.15	–0.57
10	2.78	–5.41	–9.52	–13.62	–5.65	–7.24
90	11.38	–4.84	–11.27	–17.71	–4.65	–5.69
Not applicable	1.27	3.57	–4.24	–12.05	1.64	–5.56
Typical place of service						
ED	1.58	92.79	82.02	71.25	89.49	49.30
Inpatient	8.93	–2.30	–6.49	–10.67	–2.63	–3.60
Office	1.75	10.65	5.12	–0.42	10.68	4.27
Ambulatory facility (Outpatient or ASC)	4.68	–9.05	–17.35	–25.66	–9.33	–9.45
Risk category						
Office-based	1.47	20.80	16.80	12.80	20.71	12.99
ASC	4.49	–8.26	–16.08	–23.90	–8.68	–9.15
Hospital outpatient	7.20	–3.07	–8.68	–14.29	–6.35	–5.88
Inpatient only	22.60	–1.51	–5.79	–10.07	–0.27	–3.23
Body system grouping						
Nervous system	2.53	2.27	–2.70	–7.68	0.24	–1.48

	CMS Estimate Mean RVUs	Model 1a % Difference	Model 1b % Difference	Model 1c % Difference	Model 2 % Difference	Model 3 % Difference
Endocrine system	11.79	–4.50	–13.75	–23.00	–5.63	–5.91
Eye	4.96	–0.01	–6.61	–13.20	0.17	–2.15
Ear	3.35	5.98	0.63	–4.71	5.04	–3.83
Nose, mouth, and pharynx	2.25	–6.60	–12.55	–18.50	–8.56	–8.54
Respiratory system	2.83	–4.39	–7.13	–9.88	–6.11	–4.78
Cardiovascular system	6.45	–0.48	–4.14	–7.80	–2.31	–2.24
Hemic and lymphatic system	8.14	–0.34	–4.36	–8.38	–0.71	–0.86
Digestive system	4.83	–15.92	–24.19	–32.47	–16.55	–14.21
Urinary system	3.15	–3.35	–10.61	–17.87	–3.56	–2.73
Male genital organs	7.20	–11.85	–19.76	–27.67	–8.48	–7.88
Female genital organs	5.39	–0.53	–7.19	–13.84	–1.38	–1.53
Musculoskeletal system	3.54	6.80	0.69	–5.41	7.51	1.05
Integumentary system	2.98	4.39	–3.39	–11.17	6.10	–1.47
Miscellaneous services	1.14	–3.02	–10.15	–17.28	–3.52	–7.39
Number of annual Medicare procedures						
Less than 1,000	12.29	1.28	–3.93	–9.14	1.86	0.47
1,000 to 9,999	8.87	–1.71	–6.99	–12.28	–2.47	–2.01
10,000 to 99,999	5.38	–4.15	–10.30	–16.44	–4.37	–6.45
100,000 or more	3.16	0.61	–6.24	–13.09	0.43	–2.83

Although the percentage differences vary across the models, the general patterns hold across models. For example, procedures taking 0 to 30 minutes have an average increase across all models while on average the other time categories have reductions. Across the other time categories, the smallest percentage reductions are in the procedures taking more than 120 minutes. The shorter procedures tend to have 0-day global periods and are often performed in office settings, and except for 0-day global periods in Model 3, these two categories (office-based and 0-day globals) also increase in Models 1a and 3. There are only a handful of procedures where the typical setting is ED, but the increases in these procedures are notable and warrant clinical review. Procedures typically performed in ambulatory facility settings have larger decreases than those performed as inpatient procedures, including in Model 2 (all POS). Across the body system groupings, a few body systems have increases in Models 1a and 2 (e.g., ear and musculoskeletal systems), but most systems have on average reductions in total work in the remaining models, including Model 3. The largest decreases are in procedures for the digestive system, nose, mouth, and pharynx, and male genital organ procedures. The percentage reductions, however, can be quite different across the models for these procedures. For example, the difference between CMS and predicted RVUs for digestive system procedures ranges from –14.2 percent in Model 3 to –32.5 percent in Model 1c. The latter impact is consistent with the –40 percent difference between the RAND time estimate and the CMS estimate (Table 4.2).

Figure 8.4. Distribution of Total Work RVUs

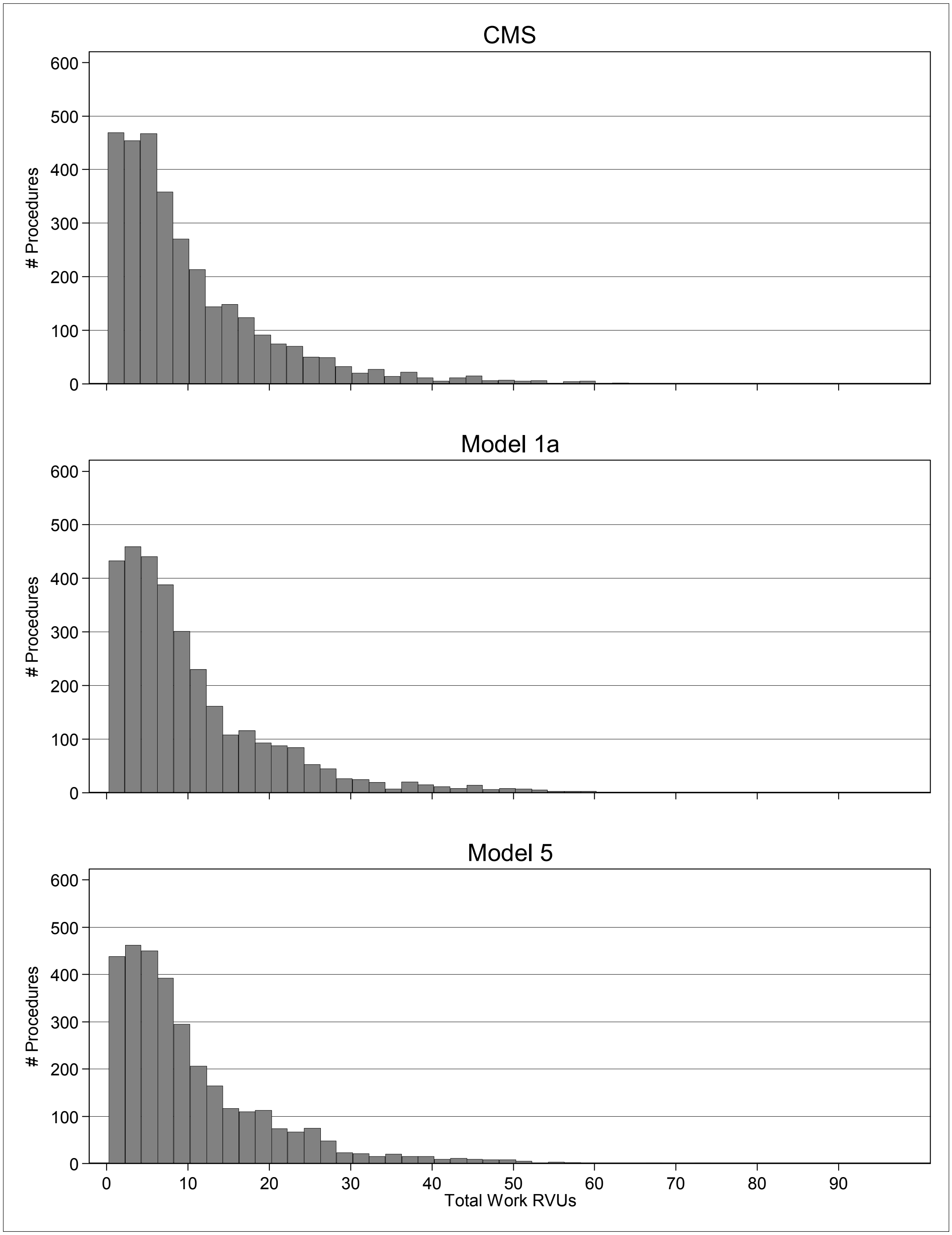


Table 8.4 reports the percentage change in the total work RVUs for selected high-expenditure procedures. The table highlights that while the overall change across models may be similar, the differences at the procedure-code level may be quite different. Most of the time the direction is the same, but the magnitude of the change is different across the models. But this is not always the case. For CPT 11721, debridement of 6 or more nails, the change ranges from –5.7 percent in Model 3 to +15.6 percent in Model 1a. This procedure is atypical in that the RAND intra-service procedure time is 28 percent longer than the CMS estimate.

Table 8.4. Percentage Difference Between CMS Total Work RVUs and RAND Estimates, "Top 20" Codes

Code	Description	CMS Total Work RVUs	Model 1a % Difference	Model 1b % Difference	Model 1c % Difference	Model 2 % Difference	Model 3 % Difference
13132	Complex laceration repair	4.8	–26.1	–31.2	–36.3	–26.0	–19.0
17311	Mohs procedure of skin	6.2	–4.0	–9.9	–15.7	–2.6	–8.2
20610	Drain or inject a bursa or joint	0.8	63.0	56.3	49.7	63.5	48.3
27245	Treat thigh fracture	18.2	–13.9	–19.6	–25.3	–10.8	–22.3
27447	Total knee arthroplasty	20.7	–4.2	–10.2	–16.2	–0.6	–8.8
31231	Diagnostic nasal endoscopy	1.1	–21.3	–24.0	–26.7	–22.5	–17.3
33533	CABG arterial single	33.8	11.3	13.7	16.2	11.3	10.5
35301	Rechanneling of artery	21.2	–3.9	–8.9	–13.8	–5.9	–0.8
43239	Esophagogastroduodeno-scopy (EGD) with biopsy	2.5	–29.4	–35.0	–40.6	–31.6	–31.3
44120	Removal of small intestine	20.8	–2.2	–6.9	–11.5	–1.2	–0.6
45380	Colonoscopy with biopsy	4.4	–26.4	–39.9	–53.4	–27.8	–21.8
47562	Laparoscopic cholecystectomy	10.5	–15.6	–22.7	–29.7	–13.8	–7.5
52000	Cystoscopy	2.2	–14.6	–21.4	–28.2	–13.9	–15.2
52601	Prostatectomy (TURP)	15.3	–28.7	–37.2	–45.6	–17.0	–20.3
62311	Inject spine lumbar/sacral	1.2	23.9	18.2	12.6	17.2	12.4
63047	Remove spinal lamina (lumbar)	15.4	7.3	2.4	–2.5	9.7	7.0
64450	Digital nerve block	0.8	51.5	46.7	41.9	39.1	54.7

Code	Description	CMS Total Work RVUs	Model 1a % Difference	Model 1b % Difference	Model 1c % Difference	Model 2 % Difference	Model 3 % Difference
66984	Cataract surgery with intraocular lens (1 stage)	8.5	–18.2	–26.2	–34.1	–18.0	–16.5
67228	Treatment of retinal lesion	13.8	–34.3	–39.7	–45.2	–35.4	–35.2
93458	Left heart artery and ventricle angiography	5.8	–9.2	–13.0	–16.8	–20.4	–9.5

In Table 8.5, we show the aggregate difference between the CMS and RAND estimates for the specialties with the highest total work RVUs for the core procedures. Both the total work RVUs attributable to all services provided by each specialty in 2012 and to the core procedures are shown. Specialties with the largest percentage reductions in RVUs for the core procedures in Model 1a are gastroenterology (–24.4 percent), colon and rectal surgery (–8.6 percent), dermatology (–7.9 percent), and urology (–6.1 percent). These specialties also have among the largest reductions in the remaining models. The specialties with the largest percentage increases in Model 1 are emergency medicine (39.6 percent), family practice (32.8 percent), and podiatry (24.1 percent).

Table 8.5. Percentage Difference Between CMS Total RVUs and RAND Estimates for Core Procedures, by Specialties with High Total Work RVUs for the Core Procedures (in Descending Order of Total Work RVUs in Core Procedures)

Specialty	Total Work RVUs for All Services (000s)	Total Work RVUs for Core Procedures (000s)	Model 1a % Difference	Model 1b % Difference	Model 1c % Difference	Model 2 % Difference	Model 3 % Difference
All specialties	1,190,007,252	191,333,481	–1.4	–7.7	–14.0	–1.7	–3.9
Ophthalmology	65,115,349	28,964,460	–0.3	–6.9	–13.6	–0.1	–2.5
Orthopedic surgery	47,339,408	28,356,192	3.4	–2.7	–8.8	4.5	–1.2
General surgery	33,480,446	20,689,787	–6.4	–12.5	–18.6	–6.0	–5.3
Gastroenterology	29,332,155	15,763,986	–24.4	–34.6	–44.8	–25.4	–22.6
Urology	23,941,648	10,737,233	–6.1	–13.7	–21.4	–5.7	–5.6
Cardiology	86,986,589	9,834,686	–3.8	–8.6	–13.3	–8.3	–5.4
Dermatology	30,445,555	8,878,978	–7.9	–14.6	–21.3	–7.9	–10.7
Vascular surgery	10,003,548	6,103,079	–0.9	–5.8	–10.7	–1.2	–1.0
Radiology	62,278,335	5,699,906	3.7	0.8	–2.0	3.0	–1.4
Neurosurgery	10,380,410	5,623,774	–5.4	–10.6	–15.7	–5.7	–7.5
Otolarngology	12,881,066	4,698,877	–6.7	–12.7	–18.6	–7.8	–8.9
Podiatry	23,834,768	4,351,163	24.1	15.2	6.3	25.5	14.6
Cardiac surgery	5,788,987	4,206,099	5.2	5.0	4.8	5.2	3.8
Thoracic surgery	5,400,573	3,902,117	4.4	3.3	2.2	4.4	3.1
Anesthesiology	7,969,118	3,790,135	–1.6	–5.0	–8.4	–3.9	–3.7
Plastic surgery	4,856,216	3,414,567	–5.8	–14.4	–23.0	–5.1	–8.9

Obstetrics/gynecology	9,139,178	3,259,292	–2.4	–9.0	–15.6	–3.1	–3.4
Internal medicine	182,827,710	3,115,331	4.1	–4.3	–12.8	3.3	–0.1
Interventional pain management	7,615,536	2,700,991	0.5	–4.3	–9.1	–3.1	–8.0
Emergency medicine	62,314,636	2,062,138	39.6	31.4	23.3	38.9	20.5
Family practice	92,714,716	2,034,956	32.8	24.8	16.8	33.9	23.0
Colon and rectal surgery	2,296,985	1,664,160	–8.6	–16.1	–23.6	–8.4	–7.7
Physician assistant	18,560,456	1,493,614	32.5	27.3	22.1	32.9	23.3
Physical medicine	15,129,132	1,448,352	7.8	2.7	–2.4	4.7	–0.7

[1] The difference is shown as a percentage of work RVUs in the core procedures. The core procedures account for varying percentages of the total work RVUs for all procedures performed by the specialty. To estimate the percentage change in total work RVUs, multiply the percentage of total work RVUs accounted for by the work RVUs for the core procedures (total work RVUs in core procedures ÷ total work RVUs – all services) by the percentage change for the model. For example, for all specialties, Model 3 results in a 3.9 percent reduction in total work RVUs for core procedures but a 0.6 percent reduction in total work (–0.039 x 191,333,481 ÷ 1,190,007,252 = –0.0063).

Relativity Under the Model Options

Spearman's rank-order correlation coefficient describes whether the relationship between two variables is monotonic. We use this measure to assess the extent to which the rank order of the total work RVUs have changed overall and within categories. Overall, the Spearman rank correlations are high and similar across the models (Table 8.6). In general, we would expect the within-category correlations to be lower than the overall value. Categories with the most homogenous work values would be expected to have the lowest correlations. Consistent with this expectation, the time categories have some of the lowest correlations, and the coefficients are particularly low (0.73–0.75) in the shortest time category. The coefficients within the global period categories are highest for the procedures with 90-day global period procedures and for inpatient procedures. Across the body system groupings, the coefficients are high, which indicates that despite changes in the RVUs for individual codes, the rank order of the values remained fairly consistent.

Table 8.6. Comparison of Spearman's Rank Order Correlation Coefficients for Core Procedures Across Models, by Procedure Category

	Medicare Volume ($ millions)	Number of Codes	Model 1a Coefficient	Model 1b Coefficient	Model 1c Coefficient	Model 2 Coefficient	Model 3 Coefficient
All codes	62.7	3,179	0.96	0.95	0.96	0.97	0.96
CMS intra-service time categories							
0 to 30 minutes	46.2	576	0.73	0.73	0.73	0.75	0.74
31 to 70 minutes	12.0	1,325	0.84	0.82	0.83	0.86	0.81
71 to 120 minutes	3.4	714	0.80	0.75	0.76	0.83	0.78
Over 120 minutes	1.1	589	0.88	0.86	0.88	0.89	0.88
Global period							
0	46.8	677	0.88	0.85	0.86	0.89	0.90
10	5.5	373	0.88	0.86	0.87	0.89	0.85
90	9.0	2,118	0.94	0.93	0.93	0.95	0.93
Not applicable	1.4	36	0.95	0.91	0.93	0.94	0.93
Typical place of service							
ED	1.1	66	0.81	0.82	0.80	0.81	0.82
Inpatient	7.5	1,290	0.96	0.95	0.96	0.96	0.95
Office	36.9	558	0.91	0.90	0.90	0.91	0.91
Ambulatory facility (Outpatient or ASC)	17.2	1,290	0.91	0.88	0.90	0.93	0.90

	Medicare Volume ($ millions)	Number of Codes	Model 1a Coefficient	Model 1b Coefficient	Model 1c Coefficient	Model 2 Coefficient	Model 3 Coefficient
Risk category							
Office-based	35.5	446	0.85	0.85	0.84	0.86	0.88
ASC OPPS	23.4	1,917	0.92	0.90	0.91	0.93	0.91
Hospital outpatient	2.1	161	0.88	0.85	0.87	0.91	0.89
Inpatient only	1.7	680	0.90	0.88	0.89	0.91	0.90
Body system grouping							
Nervous system	5.3	236	0.98	0.97	0.97	0.98	0.98
Endocrine system	0.1	25	0.87	0.82	0.82	0.89	0.88
Eye	6.2	204	0.90	0.88	0.90	0.91	0.91
Ear	1.7	41	0.95	0.94	0.95	0.94	0.93
Nose, mouth, and pharynx	0.9	141	0.96	0.96	0.96	0.97	0.95
Respiratory system	1.5	122	0.95	0.95	0.95	0.95	0.96
Cardiovascular system	5.1	353	0.95	0.94	0.94	0.96	0.95
Hemic and lymphatic system	0.1	29	0.93	0.92	0.92	0.92	0.92
Digestive system	6.8	427	0.97	0.96	0.96	0.98	0.97
Urinary system	2.6	164	0.97	0.96	0.96	0.97	0.97
Male genital organs	0.4	75	0.97	0.96	0.96	0.98	0.97
Female genital organs	0.5	132	0.97	0.96	0.97	0.97	0.97
Musculoskeletal system	10.3	891	0.94	0.92	0.93	0.95	0.92
Integumentary system	17.9	279	0.93	0.91	0.92	0.94	0.93
Miscellaneous services	3.3	85	0.82	0.82	0.79	0.81	0.86
Number of annual Medicare procedures							
Less than 1,000	0.6	1,652	0.96	0.95	0.95	0.97	0.95
1,000 to 9,999	3.5	1,059	0.96	0.95	0.95	0.97	0.96
10,000 to 99,999	13.4	410	0.94	0.93	0.93	0.95	0.95
100,000 or more	45.1	83	0.86	0.84	0.83	0.87	0.89

Summary

This chapter presents and compares the results for different options for validating the work RVUs for surgical procedures. Several models produce similar overall results, but there are significant differences for individual procedures and some systematic differences across different types of procedures. Each option has a theoretical underpinning, and the results do not suggest that one model is clearly better than another. In predicting the various work components in

Models 1 and 2, we have "smoothed" out differences in the values that could either stem from procedure characteristics that are not accounted for in our models or are anomalies in the CMS estimates. Clinical review is needed to determine whether, for example, the relative increases in pre- and immediate post-operative times for high volume office-based procedures are appropriate or whether some refinements in the model might be appropriate.

9. Other Issues

Overview

In this chapter, we discuss selected topics that are not explicitly addressed in our models but that could potentially affect how the work RVUs are estimated and/or implemented in the Medicare physician fee schedule. The RAND time estimates are based on circumstances when only a single procedure was performed; in this chapter, we first discuss our estimates of the incremental time associated with performing multiple procedures or add-on procedures to a base code. Next in this chapter we explore whether the small percentage of Medicare beneficiaries who have a core procedure as an outpatient and are subsequently admitted as an inpatient might affect how we identify the typical setting and the corrections that we make to the CMS estimates for E&M visits. Lastly, we discuss two topics that arose during the 2015 rulemaking process that affect how surgical procedures are valued. First, CMS announced in the final rule that the 90-day and 10-day global periods will be phased out and replaced with 0-day global periods. Second, CMS solicited comment on how to value codes for which moderate (conscious) sedation administered by a surgeon is an inherent part of the procedure. CMS chose not to make any immediate changes to procedures where conscious sedation is provided but indicated that it will address this in future rulemaking. We explore how our models could be refined to generate adjusted work RVU predictions that would be consistent with changes in current policies.

Multiple Procedures

Under current payment policy, the "multiple surgery rule" applies when sets of services are performed on the same patient, on the same day, and by the same provider. The multiple surgery rule accounts for the fact that there are efficiencies of scale in such situations: The provider only needs to scrub in once, only needs to make an introduction to the patient once, and may perhaps make only a single incision to access a particular organ. For a subset of surgical services, the policy values the highest work RVU procedure at 100 percent, the second-highest RVU procedure at 50 percent, and the third through the fifth highest-valued procedures at 25 percent of the fee schedule amount. Although we do not have data to address changes in all portions of work when multiple procedures are performed simultaneously, we do have a large amount of data to address the question of efficiencies in intra-service time.

The details of our methodology are provided in Appendix F. We focus solely on Medicare anesthesia times that correspond to exactly two surgical procedures and where the 50 percent multiple surgery rule applies in current CMS policy. We seek to find an appropriate multiplier p such that the time for the pair of services is well-estimated by the time estimate for the longer procedure plus p times the intra-service time estimate for the shorter procedure. We will refer to the factor applied to the shorter service as the "second code multiplier." Mathematically, our model is

$$\log(\text{time for pair}) = \log(\text{typical primary time} + p\,(\text{typical secondary time})) + \varepsilon$$

where ε is a normally distributed error term. See Appendix F for details.

Across all procedures, we estimate the second multiplier or p at 17.2 percent. We also estimate the second multiplier separately based on the organ system (as indicated by CCS level 1 groupings) of the primary (longer) service (Table 9.1). There is a substantial amount of variation in the estimates across the level 1 groupings. For two body systems (nervous and endocrine systems), we actually estimate negative second code multipliers; these are denoted with "0" in the Estimated p column. Although we do not believe that adding a procedure adds no time to the encounter, this is evidence that the addition is small on average. On the other end of the spectrum, for procedures related to the digestive system, the estimated multiplier is 37.0 percent. This category also has the highest volume in our analysis.

Table 9.1. Estimated Percentage of Intra-service Time (p) Required to Perform a Second Procedure

CCS Level 1	Description	Estimated *p*	Lower 95%	Upper 95%	N
	All	17.2%	17.0%	17.4%	750,167
1	Nervous system	7.0%	6.0%	8.0%	25,116
2	Endocrine system	0	-	-	3,643
3	Eye	0	-	-	58,160
4	Ear	4.8%	0.7%	9.1%	2,274
5	Nose, mouth, and pharynx	12.4%	10.5%	14.4%	10,115
6	Respiratory system	30.0%	28.4%	31.6%	23,562
7	Cardiovascular system	1.4%	0.8%	2.0%	93,034
8	Hemic and lymphatic system	35.0%	32.6%	37.3%	6,419
9	Digestive system	37.0%	36.5%	37.4%	239,860
10	Urinary system	29.5%	28.8%	30.1%	74,315
11	Male genital organs	19.4%	18.3%	20.6%	23,888
12	Female genital organs	21.3%	20.0%	22.6%	20,284

CCS Level 1	Description	Estimated *p*	Lower 95%	Upper 95%	N
14	Musculoskeletal system	6.9%	6.4%	7.4%	107,283
15	Integumentary system	23.0%	22.2%	23.4%	59,550
16	Miscellaneous services	1.3%	0	4.9%	2,664

NOTE: CCS Group 13 for obstetrical procedures does not include core procedures for this study.

In order to better understand the heterogeneity in the estimated second service multipliers, we conducted sub-analyses based on the estimated lengths of the services that comprise each pair. When the service with the longer estimated time is short (primary service time less than 10 minutes), we find that the estimated multiplier tends to be higher ($p = 50$ percent). For longer services, however, the estimated multiplier decreases substantially. When primary service time is between 30 and 45 minutes, p is 16.7 percent. For primary services that are estimated to be longer than 45 minutes, the estimated p is 11.2 percent. A broadly similar pattern holds when the models are stratified by the length of the secondary service.

We also examine trends defined by the relative lengths of the primary and secondary services. Here we see that only when the secondary service is nearly as long as the primary service (say, at least 50 percent of the primary service) does the secondary service seem to add substantially to the intra-service time estimated for the pair. And, even when the two services are nearly of equal length, we estimate a second service multiplier of only 20.5 percent.

Table 9.2. Estimated Multipliers by Length of Primary and Secondary Services

Subset	Estimated *p*	Lower 95%	Upper 95%	N
Primary < 10 min	43.1%	33.4%	53.6%	414
Primary 10–20	46.6%	46.1%	47.1%	175,545
Primary 20–30	0	N/A	N/A	77,642
Primary 30–45	16.7%	16.2%	17.1%	177,174
Primary 45–60	1.5%	1.0%	2.0%	111,681
Primary > 60	11.2%	10.8 %	11.6%	207,711

Subset	Estimated *p*	Lower 95%	Upper 95%	N
Secondary < 10	34.6%	33.5%	35.7%	44,206
Secondary 10–20	27.5%	27.0%	27.9%	260,666
Secondary 20–30	13.4%	12.9%	13.9%	159,981
Secondary 30–45	8.9%	8.5%	9.4%	147,888
Secondary 45–60	11.7%	11.1%	12.3%	60,168
Secondary > 60	12.5%	12.0%	13.0%	77,258
S/P ratio < 0.1	7.2%	0	28.9%	3,801
0.1 < S/P ratio < 0.25	2.1%	0	5.8%	27,675
0.25 < S/P ratio < 0.5	9.4%	8.5%	10.2%	122,594
0.5 < S/P ratio < 0.75	22.2%	21.9	22.6%	348,549
0.75 < S/P ratio < 0.9	10.5%	10.2%	10.9%	166,841
S/P ratio > 0.9	20.5%	20.0%	21.0%	80,698

NOTE: S/P ratio refers to the ratio of the estimated secondary to primary service intra service times.

Multiple Service Adjustments via Medians

In the discussions above, the multiplier p is estimated through a nonlinear least squares regression method: We find the p such that the log time of the joint service is well-approximated by the log of the sum of the two individual services' estimated times, where the secondary service's time is discounted by the factor p. An alternative approach is to consider the time estimates to be median intra-service time estimates. Using the definition of a median, we expect approximately half of the observed times to be above the estimate, and half below. In this way, we can select a multiplier p that achieves this split. A benefit of this approach is that outlying observations do not have overly strong influence on the estimates of p.

Multipliers that are estimated via median are given in Table 9.3. These are generally rather close to the values that are estimated using least squares. Across all CCS groupings, the second service multiplier is estimated to be 17.5 percent (compared to 17.2 percent estimated using least squares). As before, CCS groupings 2 and 3 have estimates that are close to zero. Similarly, CCS groupings 8 and 9 have the highest estimates (34.8 percent and 38.8 percent, respectively, compared to 35.0 percent and 37.0 percent for the least squares estimates). Given the broad similarities between the values of the multiplier that are estimated using these two methods, we have greater confidence that the main results are not overly sensitive to the assumed normal distribution in the model of our primary estimates.

Table 9.3. Secondary Service Multipliers Estimated Using the Median Method, by CCS Groupings

CCS	All	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16
p (%)	17.5	2.8	0.8	0.0	0.8	4.5	29.2	3.5	34.8	38.8	22.8	17.9	20.8	7.0	18.2	0.7

In summary, we believe there are two key findings from our analyses on secondary service multipliers. First, we see that there is substantial heterogeneity in the estimated secondary service multiplier. Arguably, this can be interpreted as a need to value more services jointly rather than as individual services. Such heterogeneity is perhaps to be expected as certain pairs of services will entail differing levels of overlap, and therefore differing efficiencies of scale. Second, if the "single multiplier" approach to discounting multiple services is adopted, these analyses suggest that the current 50 percent secondary service multiplier is too generous in most cases, particularly when both services are long when performed by themselves.

These analyses suggest that it would be sensible to maintain the 50 percent reduction only for shorter services (e.g., pairs for which the primary service is under 20 minutes) with the multiplier decreasing smoothly (as a function of the intra-service time for the primary service) to perhaps 25 percent for longer services.

One main caveat for interpreting these results is that the current 50 percent reduction applies to work values, not time estimates. It is possible that performing multiple services during one surgical encounter results in increased intensity that would not be captured by this time-only analysis. Secondly, the surgical times used in this analysis are estimated from Medicare anesthesia times. In estimating the RAND transformation, it is possible that some pairs of services have different relationships between anesthesia and surgical times. The methods outlined here could be applied to data where the surgical time is observed directly. For example, our results could be confirmed through chart abstraction or data sources with more exhaustive records of surgical times such as NSCIP and NSAS.

Add-On Procedures

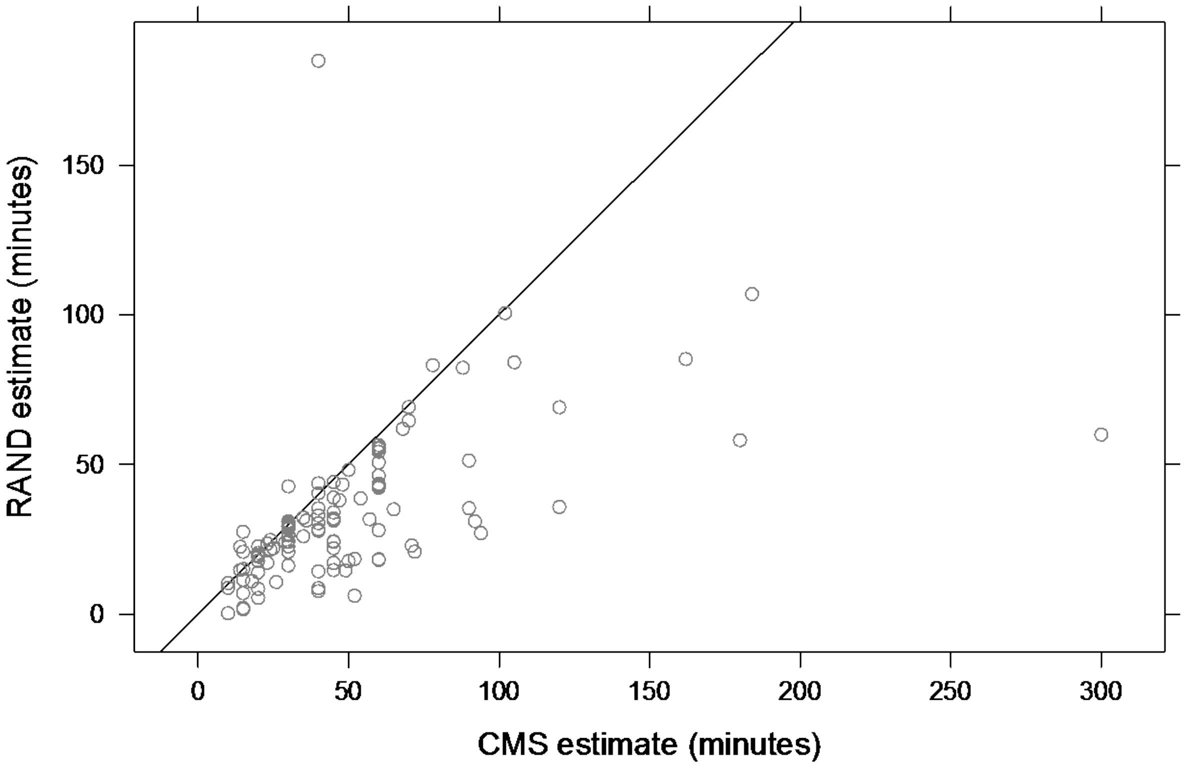
Some procedures—called add-ons—are not expected to be performed on their own and thus are valued only in conjunction with other services. For example, the CPT coding system uses primary procedure codes to report single-level spinal fusions and add-on codes to report each additional level that is fused. In this section, we discuss methods and results for updating the intra-service times associated with the add-on procedures. With these procedures, our goal is to estimate the additional time required to perform the base procedure and the add-on, relative to performing the base procedure alone (for typical cases).

Our methodology for producing these estimates is very similar to the models that we use to estimate the second code multiplier in the preceding section. Namely, we find the add-on time estimate such that the log observed times for the base/add-on pairs are well approximated by the log of the sum of estimated time for the base procedure and the (new) time estimate for the add-on. The two key differences are that the estimation is performed separately for each add-on service and that the existing CMS estimates for the add-on service carry the weight of 30 empirical observations via a Bayesian prior. See Appendix F for more details.

Our estimation results are presented in Figure 9.1. These are limited to services for which we have at least 30 anesthesia time observations of surgical encounters that comprise a single core code and a single add-on service. There are substantial differences between the CMS and RAND time estimates for several of the services. With one large exception, the RAND estimates tend to be shorter than the CMS estimates.

The services for which our estimates deviate from the CMS value by more than 60 minutes are presented in Table 9.4. Several features of this table merit mention. The "Pct CMS Estimates" column gives the percent of time observations for which the sum of the CMS time estimates for the base procedure plus the add-on service is longer than the observed surgical time (i.e., the billed anesthesia time after application of the RAND transformation). Recall that when time estimates are appropriate for the "typical" case, we anticipate that approximately 50 percent of the observed times will be longer than the estimate, and 50 percent will be shorter. However, for all of the displayed codes, the proportion of observations where the observed time is shorter than the estimate is approximately 90 percent or higher, with the exception of CPT 93462 (left heart catheterization), which is only 5.2 percent. (Because some add-on codes have multiple base procedures and because the time associated with the add-on should apply to any of the base procedures, we stratify the analysis on the basis of the add-on, rather than the combination of add-on and base procedure.) In contrast, this proportion is uniformly closer to the desired 50/50 split for the RAND estimates. With the exception of CPT 33225, the last two columns of the table are alike in being either above or below 50 percent; this is at least partially driven by the influence of the Bayesian estimation process that gives the CMS estimates substantial weight in the analysis.

Figure 9.1. Comparison of Existing CMS Estimates and RAND Estimates for Add-On Procedure Times



NOTE: Results are limited to add-ons for which we have at least 30 observations in the Medicare anesthesia analytic file.

Table 9.4. Comparison of CMS and RAND Times Where There Is More Than a 60-Minute Difference Between the Two Estimates

Add-on CPT Code	CMS Time	RAND Time	Difference	% of CMS Estimates (Base Plus Add-on) Longer Than Observed Times	% of RAND Estimates (Base Plus Add-on) Longer Than Observed Times
35700	300	60.0	240	99.0%	74.2%
93462	40	184.9	–144.9	5.2%	43.4%
35390	180	58.1	121.9	97.6%	75.9%
33225	120	35.8	84.2	94.3%	41.2%

Add-on CPT Code	CMS Time	RAND Time	Difference	% of CMS Estimates (Base Plus Add-on) Longer Than Observed Times	% of RAND Estimates (Base Plus Add-on) Longer Than Observed Times
35681	184	106.9	77.1	89.9%	76.7%
43635	162	85.2	76.8	100%	91.9%
26125	94	27.0	67.0	99.2%	60.8%
49905	92	31.1	60.9	90.8%	64.0%

NOTE: The "% of CMS Estimates" column gives the percent of observations for which the CMS time for the base procedure plus the time for the add-on code is longer than the observed surgical time. The "% of RAND Estimates" is the analogous measure with RAND (All POS) times are used for the base procedure along with the new RAND estimate of the add-on time.

Definition of "Typical" Setting and Inpatient Days

Our estimates for Models 1 and 3 are based on the "typical" setting in which a surgical procedure is performed. We use the typical setting as an explanatory variable in the prediction models and use the inpatient length of stay included in the surgery global period to validate the number of E&M inpatient visits in the CMS time estimates. In this section, we discuss our handling of outpatient surgeries that subsequently result in an inpatient admission. We explore two questions:

1. Is the definition of "typical" sensitive to how outpatient surgeries that result in a subsequent inpatient admission are classified? Our model uses the place of service reported on the physician's bill to determine where a surgical procedure is performed. If the physician reports that a procedure was performed in a hospital outpatient setting, we classify the procedure as a hospital outpatient surgery even if the patient was subsequently admitted. We then compile procedure counts by setting to validate the typical setting and re-designate the typical setting for some procedures from an inpatient to outpatient setting based on our billing data. The question is whether this re-designation would also occur if outpatient procedures that result in a subsequent admission were counted as an inpatient procedure in determining the typical setting.
2. Are the corrections that are made to the number of post-operative inpatient visits in Chapter 5 sensitive to whether to observation days occurring during the global period are included in the median inpatient length of stay measure? Medicare defines observation care as ongoing short-term treatment, assessment, and reassessment of a beneficiary before a decision is made regarding whether the patient should be admitted for inpatient care or discharged from the hospital. It often occurs when patients treated in the emergency department require significant treatment or monitoring before a decision is made regarding their admission or discharge. If a

decision is made to admit the patient, the inpatient length of stay is defined from the actual date of admission as an inpatient through discharge. In our prediction models for post-operative visits (Chapter 6), we make several corrections to the derived variables for E&M post-operative visits. Unless a procedure is typically performed in an inpatient setting, we remove any associated inpatient E&M visits in the CMS estimates. If the typical setting is inpatient, we restrict the number of inpatient post-operative visits to the number of inpatient days that are included in the global period. The count starts with the day before surgery is performed and ends with the earlier of the discharge or the global period. We make no adjustment for observation services that would be included in the global period. The question is whether the corrected count for a given procedure would be affected if the days spent in observation following an outpatient surgery were also included in the inpatient length of stay measure.

Our Medicare administrative data includes services billed by physicians and hospitals (inpatient and outpatient). We construct an analytic file that contains physician billings for core procedures with 90- or 10-day global periods with the place of service reported as ambulatory surgical center or hospital outpatient. We link the physician bills with (1) any hospital outpatient bills for the same beneficiary with the same date of service or same date minus 1 day and (2) any inpatient bills with an admission date within three days following the date of surgery. Table 9.5 reports for the core procedures the number of outpatient surgeries that are associated with an inpatient bill and the number of days between the surgery and admission. Overall, these account for less than 0.5 percent of surgical procedures with global periods.

Table 9.5. Number of Outpatient Surgeries That Result in Inpatient Admissions and Distribution of Days Between Surgery and Admission

Number of Medicare Outpatient Surgeries with Subsequent Admission	Same Day Admission	Admission 1 Day After Surgery	Admission 2 Days After Surgery	Admission 3 Days After Surgery
376,454	65.2%	16.3%	10.0%	8.5%

We found that how these outpatient procedures are defined could potentially affect the classification of 33 procedure codes. Table 9.6 lists the top 10 core procedure codes by volume that would be designated as inpatient rather than outpatient in our models if the outpatient procedures with subsequent admissions were included in the inpatient counts.

Table 9.6. Number of High-Volume Procedures Affected by Classification of Outpatient Procedures with Subsequent Admission (by Setting) and Impact on Calculation of Inpatient Percentage

CPT Code	All POS	Inpatient	Hospital Outpatient	ASC	Count of Outpatient Procedures with Subsequent Admission	Percentage Inpatient in Prediction Model	Percentage Inpatient if Outpatient Procedures with Subsequent Admission Included
19303	29,843	14,024	14,592	1,005	2,210	47.0%	54.4%
19357	7,933	3,387	3,776	610	705	42.7%	51.6%
24685	9,263	4,514	3,640	995	278	48.7%	51.7%
27385	5,447	2,631	1,984	800	261	48.3%	53.1%
28715	3,022	1,324	1,368	314	367	43.8%	56.0%
36140	31,136	15,528	11,147	34	127	49.9%	50.3%
49422	11,354	5,653	4,899	418	161	49.8%	51.2%
52601	57,317	24,367	27,524	5,023	5,011	42.5%	51.3%
93456	11,527	5,752	5,426	15	364	49.9%	53.1%
93461	24,818	12,011	11,726	22	780	48.4%	51.5%

Shifting to 0-Day Global Periods

In its final rule for CY 2015, CMS announced its intent to transition procedures with 10-day global periods to 0-day global periods in CY 2017 and procedures with 90-day global periods to 0-day global periods in CY 2018 (CMS, 2014b). The change is prompted by concerns over payment accuracy for the E&M post-operative visits included in the 10- and 90-day global periods. Implementation of this policy raises important issues concerning how to value the procedures with 0-day global periods and how to establish budget neutrality for this policy change. Implementing the 0-day global policy in a budget neutral manner requires understanding the actual pre-operative and post-operative visits that are being provided during the global period but would be separately paid under a 0-day global period. In contrast, valuing the physician work RVUs requires eliminating the physician work associated with the assumed visits that are in the current global periods.

Because we developed our models to validate the physician work RVUs under the current global period policies, the total work RVUs estimated by the model will not be relevant when CMS transitions to 0-day global periods. However, adjustments could be made in the model to estimate physician work values under a 0-day global period. With these adjustments, the model could be used to estimate the RVUs for the visits that are currently included in the total RVUs and would no longer be included when the 0-day global is adopted and subtract them from the total work RVUs. The remaining RVUs would value the procedure consistent with a 0-day global period. The results could be used to validate the physician work RVUs under a 0-day global period policy but do not inform the budget neutrality issue, which requires estimating the actual visits that would be paid separately.

Under CMS policy, the 0-day global periods include pre-operative and post-operative visits occurring on the day of the procedure. To model the 0-day global policy, we would need to consider what adjustments would be appropriate in the CMS estimates for pre-service and post-operative visits. If we were to make these adjustments, we would need to consider the following issues:

1. Pre-service visits. The 0-day global period package includes pre-service visits on the day of the procedure. The 10-day global period has the same policy so no adjustments to the pre-service visits would be needed for these procedures. The 90-day global period includes any pre-service evaluation visit the day before the procedure. The CMS time file gives a single time for pre-service review, and we would need to make assumptions regarding how much of the time is for a visit and whether the visit typically occurs the day of or day before the procedure.
2. Post-operative visits. The E&M visits for both 10-day and 90-day global periods would need to be adjusted. For a procedure typically performed in an outpatient setting, a reasonable assumption would be that any E&M visits occur after the day of surgery and would become separately billable under a 0-day global period. For a procedure typically performed in an inpatient setting, we would need to make assumptions regarding whether any of the E&M inpatient visits are likely to occur on the day of the procedure. Our current model compares the number of inpatient visits in the CMS estimates to the number of days from the date of the patient's surgery to discharge. We might assume, for example, that if the number of E&M visits in the CMS estimates is less than the median number of inpatient days, no post-operative visits occur on the day of the procedure. If the number of visits is greater, a post-operative visit occurs on the day of the procedure.

One challenge in valuing the 0-day global period is that the E&M visits may be overstated in the 90-day and 10-day global periods, leading to an overvaluation of post-operative visits and an undervaluation of intra-service work. The corrections that we make in our models, both with respect to the number of E&M inpatient visits and to the floor on intra-service intensity values, are intended to address this issue. Further, the prediction models smooth out the values consistent with those that would be expected based on intra-service time and other procedure and patient characteristics. Once the adjustments discussed above are made to the CMS estimates, the framework of our current models could be used to estimate the adjusted work RVUs using the BBM method and/or a single prediction model.

Moderate (Conscious) Sedation

Appendix G of the CPT codebook lists procedures that include moderate (conscious) sedation in the procedure description.[34] The list includes high-volume procedures, such as colonoscopies and upper GI endoscopies. Moderate sedation is not billed separately when provided by the physician who performs the procedure. However, anesthesia services provided by an anesthesiologist or, in many states, other individuals trained in anesthesia administration, such as nurse anesthetists, may be billed separately for these procedures without reducing the payment to the physician performing the diagnostic or therapeutic procedure.

Until recently, the Appendix G codes have been priced assuming that it is typical for the same physician who is performing the procedure to administer and monitor conscious sedation. . Citing changes in practice patterns and a trend away from the use of moderate sedation toward a separately billed anesthesia service, CMS solicited public comment in the CY 2015 proposed rule concerning approaches that would accurately pay for moderate sedation when it is furnished while avoiding duplicate payments when anesthesia is furnished and billed separately. Specifically, if moderate sedation were to be priced separately, this might require removing the existing valuations for moderate sedation from the work RVUs of the Appendix G procedures and establishing separate work RVUs for moderate sedation.

The request for comment on establishing a separate payment for moderate sedation raise several issues for how we handle the Appendix G procedures in our models and what adjustments would be needed if CMS were to adopt a separate payment for moderate sedation. One issue is whether our time estimates are consistent with current policies. We use two different databases for our time estimates. The SPARCS database includes procedures furnished in ambulatory facilities with and without anesthesia. We found differences in procedure lengths based on anesthesia usage and as a result, two of our POS designations are Outpatient with Anesthesia and Outpatient Without Anesthesia. With respect to the Appendix G codes, a model of log time (with random effects for CPT) indicates that observations that report anesthesia are 4.3 percent longer than observations that do not report anesthesia use. For the non–Appendix G procedures in the SPARCS database, the observations that report anesthesia use are slightly shorter (0.8 percent) than those that do not report anesthesia. Putting all procedures into a single model, the anesthesia/Appendix G code implies a difference of around 5.3 percent. Further

---

[34] The American Society of Anesthesiologists defines moderate sedation as follows (ASA, 2009):

> **Moderate Sedation/Analgesia ("Conscious Sedation")** is a drug-induced depression of consciousness during which patients respond purposefully** to verbal commands, either alone or accompanied by light tactile stimulation. No interventions are required to maintain a patent airway, and spontaneous ventilation is adequate. Cardiovascular function is usually maintained.

analysis of this finding is warranted using other external data, including the newly available NSAS for 2010 (which was not available in time for this project).[35]

A second issue is whether our prediction models are consistent with current CMS policies for valuation of Appendix G codes. To derive the dependent variables used in our prediction models, we used the reverse BBM, which means that the starting point for our valuation is the CMS estimates. In this sense, our valuation is consistent with CMS policies. However, if the CMS values are inconsistent across the Appendix G procedure codes with respect to whether conscious sedation is included in pre-service or intra-service time, our results may have some anomalies when comparing across families of codes.

A longer term issue is what adjustments might be needed in the model if CMS decides to value conscious sedation separately. The issues involved in doing so are complex and would require further analysis of an external database, such as the 2010 NSAS, which for the first time collected data by CPT code. It would require comparing the difference between Appendix G procedure codes with anesthesia to those without and making assumptions regarding how much of any differences are attributable to conscious sedation versus other factors, such as the setting in which the service is provided.

[35] We requested and were denied access to the Clinical Outcomes Research (CORI) database for gastrointestinal endoscopic procedures. We were told that the data would not be made available to our project because of data unreliability regarding whether the endoscopist's time spent on moderate sedation was included in the "scope in/out" procedure time.

10. Key Findings and Potential Applications of the RAND Models

Overview

In this chapter, we first review the key findings from the RAND validation models. We then discuss how the RAND models could be used by CMS. We provide examples of how the values could be used in the valuation process and describe the input we received from four specialty panels on procedures. Key issues we need to address are how to account for CPT coding and practice pattern changes, how to maintain relativity with other codes that were not included in our analyses, and whether our approach could be applied to other types of codes. Finally, we end with some important limitations of our analyses and future directions of this work both in the short term and long term.

Key Findings

There are five key findings from the RAND models.

1. *RAND Time Estimates Are Typically Shorter Than Current CMS Estimates*

The RAND estimates of intra-service time, which are based on data in independent datasets, are typically shorter than the current CMS estimates. As detailed in Chapter 4, for 83 percent of the procedures, the RAND time is shorter than the CMS estimates. Our finding that intra-service time is shorter than the intra-service time in CMS estimates is consistent with prior studies. This difference in time is a critical issue because intra-service time is highly correlated with total work RVUs.

2. *On Average, Total Work in RAND Models Is Similar to CMS Estimates, but There Are Important Differences for Some Groups of Procedures*

As discussed in Chapter 8, the average total work in RAND Models 1a, 2, and 3 is nearly identical to the average total work in CMS estimates. In contrast, the average RVUs for Models 1b and Model 1c are 4.8 percent and 10.0 percent lower, respectively, than the CMS average. This reflects the reductions made in intra-service work due to RAND intra-service times being shorter on average than CMS estimates.

While on average the valuations of surgical procedures are similar in several models to the CMS RVUs, there are notable differences across the types of procedures. For example, the total work RVUs for respiratory procedures in RAND Model 1a are 7.5 percent higher on average than CMS RVUs. In contrast, the average (unweighted) total work RVU estimates for digestive system procedures are similar to the current CMS RVUs (–0.04 percent). Also, for shorter procedures (0–30 minutes), the work estimates are 14.6 percent higher than CMS estimates, while for longer procedures (<120 minutes) the work estimates are 2.7 percent lower (Table H.5).

3. The Difference in Total RVUs Across RBRVS Is Greater Than the Average Difference Across Procedures for Some RAND Models

The average difference between the CMS and RAND estimates can be estimated using unweighted (average difference across all procedures) or weighted estimates (the differences for high-volume procedures have a greater effect). The difference between unweighted and weighted results is important because the weighted estimates capture what would be paid by Medicare. As discussed in Chapter 8, in RAND models 1b and 1c, the unweighted average work RVUs are higher than the weighted average RVUs. For example, the average total work RVUs under Model 1c as a percentage of CMS values are 90 percent and 86 percent (unweighted and weighted, respectively) (Figure 8.1). There is a greater reduction in the weighted results because high-Medicare-volume procedures have higher reductions on average in the intra-service work component than low-volume procedures.

4. Corrections Reduce Post-Operative E&M Visit Work

Post-operative visits on average make up 41 percent of total work RVUs among core surgical procedures we focused on in this report. For a subset of these procedures, we identified anomalies in the number of post-operative E&M visits assigned to a procedure. For example, we identified procedures for which there were inpatient E&M visits included in the global period, but the procedure is typically performed outside the hospital. Correcting for these anomalies reduced the unweighted average number of post-operative work RVUs by 10 percent.

5. The Difference Between the CMS Estimates and the RAND Estimates for IWPUT and Intra-Service Work Varies Across the Models

As discussed in Chapter 7, the implications of the shorter RAND intra-service time estimate on IWPUT and intra-service work vary under the RAND models. For example, under Model 1a the shorter intra-service time drives higher IWPUT. Under Model 1c, IWPUT stays the same, intra-service time is lower, and therefore intra-service work is reduced. This reduction in intra-service work is a critical issue because it drives the 10 percent difference between the CMS estimates and Model 1c predictions.

Potential Applications of our Model Results

We believe CMS could use the model, the individual components that go into the building block model, and the overall work RVUs we generate in two key ways to validate codes:

- CMS could use the RAND model estimates as another means of identifying potentially misvalued codes.
- CMS could use the RAND model estimates as an independent estimate of the work RVUs to consider when assessing a RUC recommendation.

Comparing RAND estimates and current CMS estimates will identify services where the valuation is inconsistent with the characteristics of the service. Using the RAND validation models as a counterpoint to the RUC process could help address many of the concerns with the current process (Table 10.1). In some cases further analysis will identify a clinical rationale for why a code is valued differently and the CMS estimate may be more appropriate. In other cases, the RAND validation model results will highlight when a code is misvalued.

We emphasize that the validation of a code for physician RVUs can be performed in many different ways. It is not clear that there is a clear "best estimate" among the models we present in this report. Given that the resource requirements of running the models is relatively low, multiple models could be run and CMS could use more than one of our model estimates for these applications. For example, Model 3 could be used to compare the CMS and RAND estimates for total work values and identify those with large discrepancies. The output for each work component from Model 1a could be used to identify which work component might be contributing to the potential differences in valuation.

Table 10.1. How RAND Validation Models Could Address Concerns with the Current System

Concern with the Current System as Described in Chapter 1	How RAND Validation Model May Address This Concern
RUC process is potentially biased.	The RAND model uses external databases to estimate characteristics of procedures and regression models to apply a consistent approach to estimate total work for each procedure.
Undervalued services are disproportionately reviewed.	The current RAND model could be used to review total work RVU values for procedure in the model on a periodic basis. It would need to be expanded to encompass low-volume and office-based surgical procedures and nonsurgical procedures.
Procedure times are too high.	Because it uses the time estimates provided in external databases, the RAND model likely provides a more accurate estimate of time.
The RUC depends on physician surveys rather than objective data.	The RAND model does not use data from physician surveys (except to the extent we use the CMS time estimates in estimating RAND times using the Bayesian approach).
Derived intra-service intensity values are sometimes negative.	Methods used in RAND validation models result in no negative values.

Concern with the Current System as Described in Chapter 1	How RAND Validation Model May Address This Concern
CMS may be overpaying for post-procedure care in the global period.	RAND validation models only partially address this issue. We make a correction when a procedure includes inpatient post-operative E&M visits, but the procedure is typically performed outside the hospital or the number of post-operative inpatient E&M visits exceeds the median length of stay.
RUC process does adequately address efficiency gains.	Because the RAND validation models can be run on a regular basis (e.g., yearly) for every procedure, efficiency gains can be incorporated by using regularly updated data on procedure times and setting.
RUC process has led to underpayment of primary care.	E&M visits make up the majority of services billed by primary care physicians, and the RAND validation models do not focus on E&M visits. Therefore this limitation is not directly addressed. However, as discussed in the section on relativity, RAND validation model results could have an indirect impact on E&M visit values.

We illustrate this process for several of the procedures that were in the list of 20 illustrative procedures. Across the 20 procedures there are many situations where the RAND model estimates are similar to the CMS estimates (for example, removal of small intestine [44120] is valued at 20.8 total work RVUs in the CMS estimate versus 20.7 in Model 3). However, there are also procedures where there are large differences. Procedures where there are discrepancies between the model estimates and CMS estimates may trigger greater scrutiny. For example, colonoscopy with biopsy (45380) is a common procedure in the Medicare population, and the total work RVUs in the RAND models are notably lower than the CMS estimate. The primary reason for this difference is intra-service time. In the external data sources, we find a colonoscopy takes 16.9 minutes[36] versus 51.5 minutes in current CMS estimates. Because intra-service time is the key variable in the RAND models, the resulting estimates for total work are much lower. For colonoscopy with biopsy, total work RVU values in the RAND Model 1c are much lower than in Model 1a (4.4 RVUs in CMS estimate, 3.3 RVUs in Model 1a, 2.1 RVUs in Model 1c). This difference between Model 1a versus 1c is driven primarily by differences in IWPUT. The IWPUT in the CMS estimate is much lower than the Model 1a estimate (0.06 IWPUT in CMS estimate vs. 0.14 IWPUT in Model 1a). In Model 1c, we estimate that the CMS IWPUT is slightly higher (0.06 IWPUT in CMS estimate versus 0.07 IWPUT in Model 1c). When CMS considers valuing this procedure, the various RAND model estimates could be used as part of that conversation. Given the lack of a gold standard for total work RVUs and IWPUT in particular, which IWPUT is most applicable is unclear. If intra-service work were valued separately in the future, this would facilitate this type of analysis.

CABG single arterial graft (33533) is a procedure where across the RAND models the total work RVU estimates are higher (for example, 33.8 RVUs in CMS estimate versus 37.6 RVUs in

[36] Intra-service time when averaged across all places it is performed. Endoscopy procedures were recently revalued by the RUC, and the CMS estimate might change in the final rule.

RAND Model 1a). Here there appear to be two drivers of the difference. First, the intra-service time in CMS estimates is *lower* than the RAND estimate (158 minutes versus 194 minutes). Second, the work RVUs for post-operative E&M visits is higher in the RAND models (15.9 work RVUs versus 17.9 work RVUs in Model 1a). The latter, post-operative E&M visits, explains the majority of the difference in total work RVUs between CMS and RAND model estimates. Key variables in the regression models for post-operative E&M visits include intra-service time, mortality rate, length of stay, and number of ICU days. Again, given a lack of gold standard for post-operative E&M visits, which estimate is more accurate is unknown. The discrepancy could be used to prompt discussion on this topic.

CMS has recently announced a policy change to phase out post-operative visits from the payment for a surgical procedure. Because the RAND models provide independent estimates of post-operative visit work, as discussed in depth in Chapter 9, another potential application of the RAND models is to help CMS value surgical procedure work after this policy change.

There are important limitations to the RAND models. For example, RAND estimates of intra-service time may not be useful select procedures and RAND models for intra-service work, and IWPUT might not capture all key clinical variables for certain procedures. (These limitations are discussed in more depth in the next section on clinical input.) Due to these limitations, we do not believe the current RAND models should replace the current valuation process. Below we highlight how the RAND models could be further refined with more research.

Findings from Clinical Panels

To inform the development of the RAND models, we obtained the input of four specialty-specific clinical panels. The objective was to obtain input on the clinical validity of the results and flag areas where the RAND models could be improved in the future. The goal was not to conduct a rigorous empirical clinical evaluation of the RAND models. Such an evaluation requires a much larger number of physicians across many specialties.

Physicians Participating in the Clinical Panels

We selected five specialties (gastroenterology, ophthalmology, dermatology, orthopedic surgery, and general surgery) that perform a large volume of procedures. Certain specialties were targeted to address certain issues. For example, we wanted gastroenterologists' input on the inclusion of conscious sedation in intra-service time for endoscopy procedures and dermatologists' on differences in time between office-based procedures and those done in the OR. A large fraction of dermatology procedures occur in the office setting and modeling primarily office-based procedures is a potential weakness of the RAND models.

For each specialty, we contacted the relevant specialty society. The American Academy of Dermatology, the American Academy of Orthopaedic Surgery, the American Gastroenterological Association, and the American Academy of Ophthalmology contacted their membership and helped RAND identify members who could participate in the clinical panel. A representative of the American College of Surgeons offered assistance in soliciting input from general surgeons, but we were not able to schedule this panel in time. We asked the specialty societies for a list of physicians who are primarily in clinical practice and who practice in different geographic regions, and clinical settings (academic, private practice). Each specialty society provided a list of approximately 15 physicians, and the RAND team then contacted those physicians independently to schedule the panel.

Each panel consisted of at least five physicians. While the invitation to participate in the panel emphasized that knowledge of the RBRVS system or RUC process was not necessary, the majority of the participants on the clinical panels were physicians within the specialty with intimate knowledge of the RUC process and many were current or prior RUC members.

Contents of Panel Discussion

The clinical input was obtained during a one-and-a-half-hour conference call. Either before the call or at the start of the call, the panelists individually conducted a pre-survey using a secure website. In this pre-survey we presented the physician a series of common procedures within his or her specialty. For each procedure we showed two choices for total work *relative to a base procedure*. The base procedure was a common procedure within the specialty. For example, a value of 2.0 for Procedure A relative to the base procedure meant that Procedure A was two times as much work. If the value was 0.8, it meant that Procedure A was 20 percent less work. One of the two choices was the current CMS estimate. The other choice was based on results from one of the RAND models.[37] The two choices were simply labeled Choice A and Choice B, and physicians were asked to choose the value that was more consistent with their clinical experience.

The results of the pre-survey were used as a starting place for the panel discussion. Panelists were presented examples of procedures on the pre-survey where across the panelists there was considerable agreement and procedures where there was significant disagreement. Panelists were asked what might explain those differences. We then presented procedures where there were large differences between the RAND model estimates and the CMS estimates for total work and work components (pre-service, intra-service time, intra-service work, post-operative E&M). We asked the panelists for their input on whether the RAND model or the CMS estimates seemed more clinically reasonable. We also presented common procedures for the specialty presented in rank order of intensity. This rank ordering was done several ways: current CMS estimates, different RAND models, and by intra-service time alone. We asked them whether a particular rank ordering of intensity made more sense clinically. Lastly, we elicited input on issues specific

[37] The RAND model results came from earlier versions of the RAND models that are similar but not identical to the final models presented in this report.

to the specialty. For example, in the clinical panel with gastroenterologists we discussed the delivery of conscious sedation and how it impacted intra-service time.

Findings from Clinical Panels

Difficulties with Comparing Procedures

The clinicians struggled with the task given to them in the pre-survey and other comparisons made during the calls. While the procedures were all common to the specialty, they represented a wide range of complexity. For example, it was difficult to compare knee arthroscopy and a knee replacement. Also complicating the comparisons were the differing global periods and the inclusion of post-operative E&M visits in some procedures but not others. A subset of physicians expressed considerable frustration with having just two choices in the pre-survey, as they felt that neither choice was consistent with their clinical experience. Physicians noted that there is considerable heterogeneity in what is performed under a given CPT code, and they had trouble cognitively "averaging" across all procedures they had performed. They tended to think of the last occasion where they performed the procedure. Some physicians subspecialized and did not perform the procedure presented on a regular basis. When the RAND model and CMS estimates had relatively small differences in magnitude, physicians felt that they were essentially the same. During the calls, many physicians expressed concern that the procedures commonly performed by their specialty were undervalued in current CMS payments. This concern might be reflected in the pre-survey results, where almost two-thirds of the time physicians selected the choice with higher relative work RVUs, regardless of whether that represented the CMS or RAND estimate.

Pre-Survey Results

Across the clinical panels and different procedures, the physicians made 318 choices between two values of total work RVUs of a procedure relative to a base procedure. One choice was based on current CMS estimates, and one was based on results from one of the RAND models. Across the procedures and physicians, 53 percent of the choices were for the CMS estimate, and 47 percent of the choices were the RAND estimate. The interpretation of the pre-survey results is unclear, given the difficulties physicians had in the task and which physicians participated in the calls.

Intra-Service Time

The clinical panels provided important insight on several key issues related to RAND's models. The databases RAND uses to estimate intra-service time do not include office-based procedures. One issue we explored during the calls was the potential bias this might introduce for intra-service time for procedures typically performed in a physician's office. The direction of the bias depended on the procedure and specialty. In some cases, physicians felt that office-based procedures were shorter than those done in an operating room because the operating room was only used in unusual circumstances. For example, anesthesia was required because the patient was mentally impaired or the case was particularly complex. An additional factor was that office-based procedures may be faster because nurses and other staff are more familiar with the procedure than staff who work in a hospital. For other procedures, clinicians argued that office-based procedures are longer. During an office-based procedure the surgeon is responsible for providing anesthesia, and because the patient is awake, the physician must frequently check in with the patient and make sure he or she is comfortable and readminister anesthetic. Together, this makes the procedure longer.

During several panels, and in particular the gastroenterology panel, there was discussion related to what is included in intra-service time versus pre-service time. For endoscopy procedures, there was disagreement on whether intra-service time should include or exclude the delivery of conscious sedation. This could impact the utility of the RAND intra-service times for valuation of these procedures (see Chapter 9 for a discussion of this issue).

The RAND method for measuring intra-service time is estimated using anesthesia time or operating room time. For a small number of surgical procedures this becomes problematic. The best example of this is a Mohs procedure (17311). This procedure is a common dermatologic procedure in which the surgeon first removes a very thin slice of a malignant lesion. The specimen is then immediately sent for preparation of the specimen for pathologic review. The surgeon reviews the specimen, and if the removed specimen has malignant cells, the surgeon removes another slice. This process can take many cycles, during which the patient can be in the waiting room or procedure room. Because of the iterative nature of this procedure, the typical definition of intra-service time ("skin-to-skin time") might be harder to apply, as for most of the intra-service time the patient is waiting. While the RAND model does provide an estimate of total work RVUs for this procedure, whether those are reasonable is less clear. Also, for some ophthalmology procedures (for example, CPT 67228), the 90-day global payment assumes that more than one procedure was performed during the global period. In contrast, the RAND intra-service time estimate focuses on the time required for a single procedure.

Intra-Service Intensity (IWPUT)

Unique aspects of procedures that might impact intensity may not be captured in the RAND models. For ophthalmology, the size of the surgical field and the risk of blindness were important variables not flagged by the panelists as important for intensity. For dermatology procedures, physicians noted that the intensity was impacted by the risk of facial cosmetic defects and the risk of not removing the entire malignancy and cancer recurrence.

When comparing the intensity of procedures, physicians tended to put the procedures into buckets with similar intensity. Panelists discussed whether the current system for a continuous scale for intensity makes sense and whether intensity should be put into discrete buckets. For example, as an illustration, one ophthalmologist put laser, intra-ocular, and extra-ocular procedures into three intensity buckets.

Better estimates of intensity and refining the RAND intra-service time estimates are key areas we discuss in the section below on future directions.

Updating the Models

To remain relevant, the validation models need to be able to account for changes in CPT coding and RVUs and in practice patterns.

Incorporating Changes in CPT Codes and RVUs into the Models

Incorporating the CY 2014 codes into our modeling effort required that we develop a method to "crosswalk" codes that had been deleted or revised since 2011 to their 2014 equivalents. Without taking this step, we would not have been able to estimate values for new codes. Our general approach is to rely on the CMS utilization file to estimate the percentage of the revised or deleted codes that would be reported under the new CPT codes. This allows us to develop weighting factors that we use to crosswalk the variables for the old codes (RAND times, procedure and patient characteristics, and intensity) into the new codes. The process that we use for the 2014 codes could be used on an ongoing basis to keep the RAND estimates current. As long as the necessary data on the procedure characteristics can be developed, the coefficients from the prediction models can be used to estimate values for new codes.

Accounting for Changes in Practice Patterns

The Bayesian intra-service time estimation methodology is particularly well-suited to such periodic updates. As new time data (e.g., a new year's worth of Medicare billing data) become available, the Bayesian paradigm allows for the existing RAND estimates to be recycled as the prior estimates. Just as the existing CMS values allow us provide reasonable time estimates even when we have little or no empirical data, the updated estimates should be reasonable even if the new tranche of data does not cover all services in all places of service. And, as with the adjusted CMS values in the Bayesian models, this approach allows for reasonable estimates of low-volume services, updating times where data are available and maintaining relativity where data are not available. Depending on the extent to which intra-service times are expected to change from year to year, it may be reasonable to give the prior estimates more weight than the CMS time estimates were given in our time analyses.

Maintaining Relativity

We chose to focus on a select set of surgical and medical procedures in this project. Our core codes account for approximately four-fifths of aggregate RVUs among all surgical procedures. One critical aspect of the current process is that codes are maintained in a relative manner. In other words, each code in the physician fee schedule has a relative value compared to all other codes.

In contrast, the RAND validation model does not focus on relativity across all codes. A code is valued based on the characteristics of code, and relativity is only applicable within the core procedures we focused on. One issue that needs to be addressed is how CMS could incorporate the RAND validation values into the larger RBRVS.

Current Process for Maintaining Relativity Within the RUC

When codes are updated, relativity is maintained in three ways: (1) within a code family, (2) within a specialty, and (3) across specialties. Codes are grouped into families to evaluate relativity. For example, in a recent RUC evaluation of a code, the upper gastrointestinal endoscopy family included 23 codes (43235 through 43260). Families have a base code that is the starting place for relative values. In the upper gastrointestinal endoscopy family, the base code is 43235 (work RVUs 2.39), and the other 22 codes in this family maintain relativity to this base code. That is, the RVUs of services within a family should be ranked progressively so that more intensive services should be assigned higher work RVUs relative to 43235. If we define relative values as work RVUs for code divided by work RVUs for 43235, then the relative values for the other 22 codes ranges from 1.08 (43241: upper endoscopy with transendoscopic intraluminal tube or catheter placement) to 3.05 (43242: upper endoscopy with transendoscopic FNA) using CMS estimates. If the work RVUs for one code in the family are updated, relativity with other codes in the family is typically maintained through adjustments to the RVUs for the other codes.

Relativity is also maintained within specialty. The work RVUs of codes are relative based on time and intensity of other services provided by physicians in the same specialty. For example, if a code for treatment of elbow fracture is updated, the expectation would be that other codes furnished by the same specialty (e.g., treatment of femur fracture) would maintain an appropriate relative value. We provide specific values to make this more concrete. Code 24650 (elbow fracture: closed treatment of radial head or neck) has work RVUs of 2.31, while code 27230 (femur closed treatment of femoral fracture, proximal end, neck) has work RVUs two and half times higher at 5.81. If this elbow fracture code's work RVUs were increased from 2.31 to 2.61, then, to maintain relativity, this femur fracture's code could be increased from 5.81 to 6.55 (6.55 / 2.61 = 2.5).

The last point of relativity is across specialties using the Multispecialty Points of Comparison (MPC) list of codes. The list is used to help judge relativity of procedures across specialties. The current MPC list consists of 316 codes, which are provided by more than one specialty. For example, code 22520 is percutaneous vertebroplasty which was performed 11,422 times among the Medicare population in 2010; 63 percent were performed by diagnostic radiologists, 10 percent were performed by interventional radiologists, and 9 percent were performed by orthopedic surgeons. Another procedure on the MPC is bronchoscopy (code 31622), which was performed 84,807 times in 2010, and 52 percent were performed by pulmonologists, 15 percent by thoracic surgeons, and 8 percent by internists. Because the MPC is so critical in the fee schedule, it is vital that the services on the MPC list be appropriately valued, and the rank order of the MPC has been described as the gold standard for placement of values.

Maintaining Relativity Using the RAND Models

As noted earlier in this chapter, CMS could use RAND's estimates of work RVUs (or components of the work RVUs) in different ways. If CMS uses RAND's values to validate a single code, CMS and the RUC could continue the same approach currently used to maintain relativity. That is, in revising the value for an individual code, maintaining relativity across the family of codes, specialties and the MPC could be considered.

If CMS would like to use RAND's estimates for multiple codes, the process of maintaining relativity is more complicated. The total work estimates differ between CMS estimates and RAND estimates for most common major surgical procedures. When RAND estimates a very different value for work RVUs (or component such intra-service time) compared to CMS values, what does that mean? The underlying issue is how work RVUs should be estimated. One perspective is that the work RVUs for services can be only estimated in a relative manner, and the other perspective is that work RVUs can be estimated accurately on a code-by-code basis without regard to the RVUs for other services. To make these issues more concrete, we start with three illustrative codes from the MPC and compare their CMS RVUs relative to 99213, which is a commonly used E&M visit code.

Table 10.2. Valuations in Current CMS Estimates of Three Codes from MPC

Code	Description	CMS Work RVU	Intra-Service Time in CMS Estimates (minutes)	CMS Work RVU Relative to RVUs for 99213
33533	CABG—arterial graft	33.8	151	34.8
26615	Open treatment of metacarpal fracture	7.1	45	7.3
99213	E&M level 3	0.97	15	1.0

In Table 10.3, we show the relative values of these three codes to 99213 E&M across two of our five models and the current CMS Estimates.

Table 10.3. Valuations in Current CMS Estimates of Three Codes from MPC

Code	Description	Work RVU in CMS Estimates	Work RVU in RAND Model 1a	Work RVU in RAND Model 3
33533	CABG—arterial graft	33.8	37.6	37.3

Code	Description	Work RVU in CMS Estimates	Work RVU in RAND Model 1a	Work RVU in RAND Model 3
26615	Open treatment of metacarpal fracture	7.1	7.8	7.7
99213	E&M level 3	0.97	-	-

Given the differing changes across the codes, it is mathematically impossible to maintain the current relativity of these two codes and maintain relativity between the individual codes and 99213.

Maintaining relativity between 33533 and 26615 may be less an issue. One could argue that the data sources RAND used reflect important and accurate data and that the regression models are applied in a constant manner across all the codes. The relative values in the MPC could be open to change. As CMS has argued in the past, given the rapid changes in medical practice, there is no reason to believe that the relativity of the MPC codes would not have changed over time and that if newer and more accurate data become available, then it is important to adapt the MPC values and the fee schedule accordingly.

The more difficult issue is how to address relativity for codes outside our scope (e.g., E&M visits) if the RAND estimates were used for many procedures. We believe CMS could pursue two options: (1) not maintaining relativity and (2) maintaining average relativity. While the RVUs for codes outside the surgical range would not change under either option, the payments for these codes could be affected by which option was pursued because of the budget neutrality adjustment and potential impact on the conversion factor.

The rationale for maintaining average relativity is that the underlying reasons why work RVUs decreased among the procedures in our core list (e.g., lower intra-service time) may be applicable for all other parts of the RBRVS. In other words, if RAND were to identify external data sources with intra-service time and develop prediction models, then, on average, work RVUs for these other services (e.g., E&M, radiology, pathology) would likely be decreased by similar amounts. The rationale for not maintaining relativity is that the underlying reasons why work RVUs decreased among the codes in our core list are unique to surgical procedures and are not generalizable to other types of codes, and these represent misvalued codes.

If CMS were to not maintain relativity, the RVUs for codes outside the surgical range would be unaffected, and if the budget neutrality adjustment were applied, this would increase the conversion factor. Because the work RVUs for the core procedure codes have decreased on average, this would mean less spending on physician work for procedures and higher spending on other codes including E&M visits. This would be an indirect mechanism to address the criticism of the current process that E&M visits are undervalued.

If CMS were to maintain average relativity, the simplest approach would be to apply an across-the-board adjustment factor to the core procedures so that the weighted average RVU for the procedures equal the weighted average CMS RVUs for the same procedures. This approach does not maintain relativity with the MPC codes but makes the code revisions budget neutral within the affected core procedures.

The tension between not maintaining versus maintaining relative accuracy might vary by the level of relativity. As discussed above, there are three levels at which relativity is maintained: within family, within specialty, and across MPCs.

Expanding RAND Validation Process to Other Codes

There are two parts of the physician fee schedule that we did not validate in our current analyses: (1) surgical procedures that we dropped from our core list and (2) nonsurgical codes. Nonsurgical codes include codes for E&M visits, anesthesia, radiology, pathology and laboratory, and most medicine codes (some were included in our models).

Our validation process focuses on data from external databases to estimate the time required to perform the service, the characteristics of physicians who perform the service and patients who receive the service, and their outcomes afterwards. Whether this approach is feasible and applicable to nonsurgical codes is something we explore below.

Expanding Model to Surgical Procedures Not on Our Core List

As detailed in Appendix A, we dropped codes for a variety of reasons. The vast majority of codes were dropped because they were billed less than 100 times for Medicare fee-for-service beneficiaries. We chose this cutoff because we have difficulty estimating both the intra-service time and other components of our models (for example, mortality rate). CMS must still place a work RVU value for these services. Given their rarity and relatively low impact on total payments paid by CMS, one option is to not create a validation estimate unless they become more common in the Medicare program (i.e., billed greater than 100 times). Valuing these rare codes would depend only on the RUC process.

An alternative approach is to use limited prediction models to estimate the individual components or total work RVUs for these procedures. We use the term *limited prediction models* because some variables (for example, mortality rate, complication rate) cannot be assessed accurately for procedures with only 20 or 30 observations. However, other characteristics of the code (for example, years of training for physicians who perform the procedure) could be used in the model.

Another set of procedures we dropped were surgical procedures almost always performed in settings where there is no operating room, such as a physician office. In the datasets currently available to use, we had little data on office-based procedures. If such data were to become available, then the RAND validation model could be expanded to these procedures. The research being conducted by the Urban Institute on the RBRVS could potentially be used in this regard.

Nonsurgical Codes

In this exploratory project, we focused our work on surgical procedures and some common medical procedures. Two larger issues will need to be addressed if CMS were to consider expanding the validation approach to nonsurgical codes: (1) availability of external data and (2) development of appropriate predictive models.

The first issue, availability of external data, is difficult but not insurmountable. The time required for certain E&M visits is a critical issue. For example, the newest versions of the National Ambulatory Medical Care Survey (data not yet released) have collected time data for specific E&M CPT codes. Similarly, individual health systems may have time and motion studies on time required for the face-to-face time for different types of E&M visits. Large radiology groups routinely collect the time required to interpret and report the results of different types of radiology procedures. These time data have been used in some research studies (Clark et al., 2013). It is possible that similar data sources are available for such procedures as pathology interpretation. One key issue that will arise is that these other data sources are often collected by private organizations (as opposed to government entities), and whether they would be willing to share the data voluntarily is unclear.

The second issue, development of appropriate predictive models, will require both theoretical and empirical evaluation. One key aspect of our project was using such variables as mortality, level of illness among patients receiving that service, and where the procedure was performed to predict the pre- and post-work associated with a service. While those could be estimated for nonsurgical codes, it is less clear that they *should* be included in a prediction model. For example, the 30-day mortality after certain types of E&M visits might be very high, but this is because the physicians are caring for a sick population and not that the E&M visit led to this higher mortality. Similarly, the level of illness of a patient may not be applicable to a work model for pathology interpretation.

Key Limitations

In each chapter, we discuss the limitations of the specific prediction models. Here we focus on the key overall limitations. Perhaps one of the most challenging aspects of our modeling activity has been the lack of external data that can be used to validate the different work components. The lack of a gold standard to estimate intra-service work has been particularly problematic, and, as discussed below, more work can be conducted to estimate intra-service intensity.

Throughout the modeling activity, we have needed to rely on CMS estimates for some elements used in the BBM approach, such as the intensity values for the pre-service and post-service components. Also, our prediction models for post-operative E&M visits were built using data in the current CMS estimates. If the CMS estimates are on average biased, then on average our model estimates are similarly biased. For example, there has been concern from the Office of the Inspector General that the number of post-operative E&M visits included in the global period is too high (Department of Health and Human Services, 2012). Because we lacked an external method to validate the number of post-operative E&M visits, our models could not address that concern. Our models can be seen as a mechanism to identify or smooth out inconsistencies in the current values, but they do not fundamentally shift the amount of work devoted to a component of the BBM, except where corrections are made to the current values (as we do for post-operative inpatient E&M visits).

For our intra-service time estimates, we do use external data. However, we combine two different data sources and transform the times reported in those data sets (OR and anesthesia) into comparable surgical time estimates. Neither database represents the full range of settings and clinical conditions under which the procedures are performed. Also, including only procedures that are performed in facility settings may over- or understate the times for procedures that are commonly performed in office settings. We address this issue by excluding surgical procedures from our core procedure set that are performed less than 5 percent of the time in a setting that may have an operating room (inpatient, ED, ASC, outpatient hospital), but we do not know whether this step is sufficient to avoid bias. For example, take a procedure that is performed 90 percent of the time in the office setting. Our intra-service time estimates come from the 10 percent of procedures that occur in a facility setting. Whether the intra-service time estimated from this limited set of cases is more accurate is unknown, though it is reassuring that, across our core set of surgical procedures, there was no relationship between (1) the fraction of procedures in the Medicare population that occurred in a facility setting and (2) the difference between RAND and CMS time estimates.

Most of the other independent information that we incorporate into our prediction models is based on Medicare administrative data. While this might be appropriate since we are validating RVUs used in the Medicare fee schedule, we recognize that the RVUs are intended to reflect all patients and indeed are used by non-Medicare payers. A potential next step would be to explore the sensitivity of the estimates to information that is more reflective of the total patient population.

Finally, the validation model reflects predictions based on statistical modeling and likely produces some results that are inconsistent with clinical experience. We provide the example of Mohs surgery in this chapter. Across the thousands of procedures included in this project, there are likely many other examples of where the RAND models may have spurious results.

Future Work to Refine and Expand the RAND Models

To refine the RAND models for surgical services, we recommend three areas where more research can be done:

1. Develop a gold standard for intra-service work for a small set of surgical services. This would address a key limitation of the RAND models that the intra-service work

estimates used to build the prediction models were derived from the current CMS estimates. We recommend that intra-service work for approximately 200 surgical services be valued using physician input in a process separate from the current RUC process. These gold standard values for intra-service work would be used to calibrate the RAND prediction models. The RAND prediction models could then be used to estimate work RVUs for the full set of surgical procedures.

2. Improve the RAND time estimates by obtaining more data on the times for procedures, in particular those done in an office setting. This would help expand the model to office-based procedures.
3. We believe that the RAND models could be improved by obtaining more clinical input. This might involve refining the procedure characteristics used in the current models and also adding new characteristics that are not currently included.

The current RAND models focus only on surgical procedures. As discussed above, research could be conducted to expand the RAND models to the nonsurgical aspects of RBRVS. The key will be to identify external datasets with intra-service times for these services.

Conclusions

In this project, RAND developed an independent method for valuing physician work RVUs for surgical procedures. Using external data, for each surgical procedure, we measured such characteristics as intra-service time, years of training among physicians who perform the procedure, and the mortality risk after the procedure. These are used in the models to estimate total work and the subcomponents of work for each procedure.

The methods for developing these models are complex, and because there is no single optimal approach, RAND generated different alternative models that reflect various major methodological decisions and tradeoffs. RAND's models address many of the concerns with the current process for valuing physician services and can help improve the RBRVS by identifying misvalued codes and serving as a potential counterpoint to RUC valuations of services. However, there are key limitations to the models. Future work incorporating clinical input and obtaining more data can help to address these limitations and further refine the RAND models.

References

AMA—*see* American Medical Association.

American Board of Medical Specialties, *ABMS Guide to Medical Specialties 2013*, Elsevier Press, 2013.

American College of Surgeons National Surgical Quality Improvement Program, home page, 2014. As of November 11, 2014: http://site.acsnsqip.org/

American Medical Association, *CPT 2013 Professional Edition*, Chicago, Ill., 2012.

American Medical Association/Specialty Society, "The RUC Relativity Assessment Workgroup and CPT," 2012.

American Medical Association/Specialty Society, "Improvements of the RUC," 2013. As of October 29, 2014: http://www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/medicare/the-resource-based-relative-value-scale/the-rvs-update-committee.page

American Medical Association, "The RUC Relativity Assessment Workgroup Progress Report," 2014a. As of October 29, 2014: http://www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/medicare/the-resource-based-relative-value-scale/the-rvs-update-committee.page

American Medical Association/Specialty Society, "RVS Update Process Booklet," 2014b. As of October 29, 2014: http://www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/medicare/the-resource-based-relative-value-scale/the-rvs-update-committee.page

American Society of Anesthesiologists, "Continum of Depth of Sedation: Definitions of General Anesthesia and Levels of Sedation/Analgesia," 2009. As of November 11, 2014: https://www.asahq.org/For-Members/Standards-Guidelines-and-Statements.aspx

Berenson, Robert A., "Out of Whack: Pricing Distortions in the Medicare Physician Fee Schedule," *Expert Voices*, National Institute for Health Care Management, September 2010. As of November 11, 2014: http://nihcm.org/pdf/NIHCM-EV-Berenson_FINAL.pdf

Bodenheimer, T., R. A. Berenson, and P. Rudolf, "The Primary Care–Specialty Income Gap: Why It Matters," *Annals of Internal Medicine*, Vol. 146, No. 4, 2007, pp. 301–306.

Bodenheimer, Thomas, Mina Matin, and Brian Yoshio Laing, "The Specialist–Generalist Income Gap: Can We Narrow It?" *Journal of General Internal Medicine*, Vol. 23, No. 9, September 2008, pp. 1539–1541.

Braun, P., J. Dernburg, D. L. Dunn, and W. Cohen, "Predicting the Work of Evaluation and Management Services," *Medical Care*, Vol. 30, No. 11, 1992, pp. NS13–NS27.

Centers for Disease Control and Prevention, "National Survey of Ambulatory Surgery," 2006. As of November 11, 2014:
http://www.cdc.gov/nchs/nsas.htm

Centers for Disease Control and Prevention, "Ambulatory Health Care Data," 2010. As of November 11, 2014:
http://www.cdc.gov/nchs/ahcd.htm

Centers for Medicare & Medicaid Services, Medicare program; Payment policies under the physician fee schedule and other revisions to Part B for CY 2010 (CMS-1413-FC), Federal Register, Vol.74, No. 226, November 25, 2009.

Centers for Medicare & Medicaid Services, Medicare program; Payment policies under the physician fee schedule, five-year review of work relative value units, clinical laboratory fee schedule: Signature on requisition, and other revisions to Part B for CY 2012 (CMS-1524-FC and CMS-1436-F), Federal Register, Vol. 76, No. 228, November 28, 2011.

Centers for Medicare & Medicaid Services, *CMS Manual System: Medicare Claims Processing*, 2012a. As of May 5, 2014:
http://www.cms.gov/Regulations-and-Guidance/Guidance/Transmittals/downloads/R2395CP.pdf

Centers for Medicare & Medicaid Services, Medicare Program; Revisions to Payment Policies Under the Physician Fee Schedule, DME Face-to-Face Encounters, Elimination of the Requirement for Termination of Non-Random Prepayment Complex Medical Review and Other Revisions to Part B for CY 2013; Final Rule (CMS-1590-FC), Federal Register, Vol. 77, No. 222, November 16, 2012b. As of November 13, 2013:
http://www.gpo.gov/fdsys/pkg/FR-2012-11-16/html/2012-26900.htm

Centers for Medicare & Medicaid Services, Medicare and Medicaid Programs: Hospital Outpatient Prospective Payment System; Ambulatory Surgical Center Payment System; Hospital Outpatient Quality Reporting Program; Electronic Health Record Incentive Program Electronic Reporting Pilot; Ambulatory Surgical Center Quality Reporting Program; Inpatient Rehabilitation Facility Quality Reporting Program; Revision to Quality Improvement Organization Regulations, Final Rule (CMS-1589-FC), Federal Register, Vol.

77, No. 221, November 15, 2012c. As of November 13, 2014: http://www.gpo.gov/fdsys/pkg/FR-2012-11-15/pdf/2012-26902.pdf

Centers for Medicare & Medicaid Services, "Table 6I-Major CC List-FY2014," Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals and the Long‐ Term Care Hospital Prospective Payment System and Fiscal Year 2014 Rates; Quality Reporting Requirements for Specific Providers; Hospital Conditions of Participation; Payment Policies Related to Patient Status (CMS-1455-F), Table 6I, Federal Register, Vol. 78, No. 160, August 19, 2013. As of March 11, 2014: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY-2014-IPPS-Final-Rule-Home-Page-Items/FY-2014-IPPS-Final-Rule-CMS-1599-F-Regulations.html?DLPage=1&DLSort=0&DLSortDir=ascending

Centers for Medicare & Medicaid Services, "Medicare Program; Revisions to Payment Policies Under the Physician Fee Schedule, Clinical Laboratory Fee Schedule, Access to Identifiable Data for the Center for Medicare and Medicaid Innovation Models & Other Revisions to Part B for CY 2015 (CMS 1612-P)," Federal Register, Vol. 79, No. 133, 2014a. As of October 24, 2014: http://www.gpo.gov/fdsys/pkg/FR-2014-07-11/pdf/2014-15948.pdf

Centers for Medicare & Medicaid Services, Medicare Program; Revisions to Payment Policies Under the Physician Fee Schedule, Clinical Laboratory Fee Schedule, Access to Identifiable Data for the Center for Medicare and Medicaid Innovation Models & Other Revisions to Part B for CY 2015 (CMS-1612-FC), Federal Register, Vol. 79, No. 219, November 13, 2014b. As of November 13, 2014: https://www.federalregister.gov/articles/2014/11/13/2014-26183/medicare-program-revisions-to-payment-policies-under-the-physician-fee-schedule-clinical-laboratory

Centers for Medicare & Medicaid Services, "Chapter 12—Physicians/Nonphysician Practitioners," *Medicare Claims Processing Manual Revision 2997*, 2014c. As of April 25, 2014: http://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/Downloads/clm104c12.pdf

Charlson, Mary E., Peter Pompei, Kathy L. Ales, and C. Ronald MacKenzie, "A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation," *Journal of Chronic Diseases*, Vol. 40, No. 5, 1987, pp. 373–383.

Clark, Jonathan R., Robert S. Huckman, and Bradley R. Staats, "Learning from Customers: Individual and Organizational Effects in Outsourced Radiological Services," *Organization Science*, Vol. 24, No. 5, September–October 2013, pp. 1539–1557.

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

Cromwell, J., S. Hoover, et al., "Validating CPT Typical Times for Medicare Office Evaluation and Management (E/M) Services," *Medical Care Research and Review*, Vol. 63, No. 2, 2006, pp. 236–255.

Cromwell, J., N. McCall, K. Dalton, and P. Braun, "Missing Productivity Gains in the Medicare Physician Fee Schedule: Where Are They?" *Medical Care Research Review*, Vol. 67, No. 6, 2010, pp. 676–693.

Department of Health and Human Services, Office of Inspector General, *Cardiovascular Global Surgery Fees Often Did Not Reflect the Number of Evaluation and Management Services Provided*, A-05-09-00054, May 2012a. As of November 11, 2014: http://oig.hhs.gov/oas/reports/region5/50900054.pdf

Department of Health and Human Services, Office of Inspector General, *Musculoskeletal Global Surgery Fees Often Did Not Reflect the Number of Evaluation and Management Services Provided*, A-05-09-00053, May 2012b. As of November 11, 2014: https://oig.hhs.gov/oas/reports/region5/50900053.pdf

Goodson, J. D., "Unintended Consequences of Resource-Based Relative Value Scale Reimbursement," *Journal of the American Medical Association*, Vol. 298, No. 19, 2007, pp. 2308–2310.

Healthcare Cost and Utilization Project (HCUP), *Clinical Classifications Software for Services and Procedures*, 2014.

Horner, R. D., J. P. Szaflarski, C. J. Jacobson, et al., "Clinical Work Intensity Among Physician Specialties: How Might We Assess It? What Do We Find?" *Medical Care*, Vol. 49, 2011a, pp. 108–113.

Horner, R. D., J. P. Szaflarski, J. Ying, et al., "Physician Work Intensity Among Medical Specialties: Emerging Evidence on Its Magnitude and Composition," *Medical Care*, Vol. 49, 2011b, pp. 1007–1011.

Hsiao, W. C., P. Braun, et al., "Results and Policy Implications of the Resource-Based Relative-Value Study," *New England Journal of Medicine*, Vol. 319, No. 13, 1988a, pp. 881–888.

Hsiao, W. C., P. Braun, et al., "Estimating Physicians' Work for a Resource-Based Relative-Value Scale," *New England Journal of Medicine*, Vol. 319, No. 13, 1988b, pp. 835–841.

Hsiao, W. C., P. Braun, D. L. Dunn, E. R. Becker, D. Yntema, D. K. Verrilli, E. Stamenovic, and S. P. Chen, "An Overview of the Development and Refinement of the Resource-Based Relative Value Scale: The Foundation for Reform of U.S. Physician Payment," *Medical Care*, Vol. 30, No. 11, 1992, pp. NS1–NS12.

Jacobs, J. P., D. M. Shahian, et al., "Successful Linking of the STS Adult Cardiac Surgery Database to CMS Medical Data to Examine the Penetration, Completeness, and

Representativeness of the STS National Database," Society of Thoracic Surgeons 2010 Annual Meeting, Fort Lauderdale, Fla., 2010.

Jacobson, C. J., S. Bolon, N. Elder N, et al., "Temporal and Subjective Work Demands in Office Based Patient Care: An Exploration of the Dimensions of Physician Work Intensity," *Medical Care*, Vol. 49, 2011, pp. 52–58.

Kahan, J. P., S. C. Morton, H. H. Farris, G. F. Kominski, and A. J. Donovan, "Panel Processes for Revising Relative Values of Physician Work, A Pilot Study," *Medical Care*, Vol. 32, No. 11, 1994, pp. 1069–1085.

Kerber, K.A., M. Raphaelson, G. L. Barkley, and J. F. Burke, "Is Physician Work in Procedure and Test Codes More Highly Valued Than That in Evaluation and Management Codes?" *Annals of Surgery*, September 2014.

Laugesen, M. J., and S. A. Glied, "Higher Fees Paid to US Physicians Drive Higher Spending for Physician Services Compared to Other Countries," *Health Affairs*, Vol. 30, No. 9, September 2011, pp. 1647–1656.

Mabry, C. D., and P. Bluff, "On the Cutting Edge of Reimbursement Strategies: The ACS Develops New Techniques," *Bulletin of the American College of Surgeons*, Vol. 85, No. 6, June 2000.

Mabry, C. D., B. C. McCann, J. A. Harris, J. Martin, J. O. Gage, J. E. Fischer, F. G. Opelka, R. Zwolak, K. Borman, J. T. Preskitt, P. E. Collicott, L. McGinnis, and I. Cohn, "The Use of Intra-Service Work per Unit of Time (IWPUT) and the Building Block Method (BBM) for the Calculation of Surgical Work," *Annals of Surgery*, Vol. 241, 2005, pp. 929–940.

McCall, N. T., J. Cromwell, E. Drozd, S. Hoover, and P. Braun, *Validation of Physician Time Data: Final Report*, Waltham, Mass.: Health Economics Research, Inc., 2001.

McCall, N., J. Cromwell, and P. Braun, "Validation of Physician Survey Estimates of Surgical Time Using Operating Room Logs," *Medical Care Research Review*, Vol. 63, No. 6, 2006, pp. 764–777.

MedPAC—*see* Medicare Payment Advisory Commission.

Medicare Payment Advisory Commission, *Report to the Congress: Medicare Payment Policy*, Washington, D.C., March 2006.

Medicare Payment Advisory Commission, *Outpatient Hospital Services Payment System*, 2007.

Medicare Payment Advisory Commission, *Report to the Congress: Medicare and the Health Care Delivery System*, Washington, D.C., June 2011a.

Medicare Payment Advisory Commission, "Moving Forward from the Sustainable Growth Rate (SGR) System, Letter to Congress, October 14, 2011b.

New York State Department of Health, “SPARCS Overview,” 2007.

New York State Department of Health, *SPARCS: X12-837 Input Data Specifications*, 2009. As of October 29, 2013: http://www.health.ny.gov/statistics/sparcs/sysdoc/input5010.pdf

Rich, William L., III, "RUC Letter to Jeffrey A. Alexander about McCall Study," February 8, 2007.

Silber, J. H., P. R. Rosenbaum, et al., "Estimating Anesthesia and Surgical Procedure Times from Medicare Anesthesia Claims," *Anesthesiology*, Vol. 106, No. 2, 2007, pp. 346–355.

Silber, J. H., P. R. Rosenbaum, et al., "Estimating Anesthesia Time Using the Medicare Claim: A Validation Study," *Anesthesiology*, Vol. 115, No. 2, 2011, pp. 322–333.

Sinsky, C. A., and D. C. Dugdale, “Medicare Payment for Cognitive vs Procedural Care: Minding the Gap,” *JAMA Internal Medicine*, Vol. 173, No. 18, October 14, 2013, pp. 1733–1737.

Smith, P. K., J. E. Mayer, Jr., K. R. Kanter, V. J. DiSesa, J. M. Levett, C. D. Wright, F. C. Nichols, and K. S. Naunheim, “Physician Payment for 2007: A Description of the Process by Which Major Changes in Validation of Cardiothoracic Surgical Procedures Occurred,” *Annals of Thoracic Surgery*, Vol. 83, 2007, pp. 12–20.

Sriram, K., R. V. Ramamoorthi, and P. Ghosh, “Posterior Consistency of Bayesian Quantile Regression Based on the Misspecified Asymmetric Laplace Density,” *Bayesian Analysis*, Vol. 8, No. 2, 2013, pp. 479–504.

Weems, K., “Letter to William L. Rich III,” Washington, D.C., Center for Medicare and Medicaid Services, June 19, 2008.

Weems, K., "Doing More with Less: A Conversation with Kerry Weems," *Health Affairs (Project Hope)*, Vol. 28, No. 4, 2009, pp. w688–w696.

Yu, Keming, and Rana A. Moyeed, "Bayesian Quantile Regression," *Statistics & Probability Letters*, Vol. 54, No. 4, 2001, pp. 437–447.

List of Appendixes

Appendix A. Defining Core Codes for Analyses

We address the following in this section: (1) explanation of why we focused on surgical codes, (2) rationale for adding select medicine codes to the surgical codes, and (3) rationale and impact of excluding certain codes.

Focus on Surgical Codes

As an illustration of the RAND building block model, we focus on the CPT category of surgical services (CPT 10021–69990). This broad category includes surgery (where there is an incision of some kind) and other types of procedures, such as colonoscopy, that do not require an incision. We chose these codes primarily because of available time data. Our goal is to validate the different components that go into calculating work RVUs for a given service. The datasets available to us for this project have data on intra-service time for surgical procedures provided in hospital inpatient and outpatient settings and in ambulatory surgical centers (ASCs). These will include more common ophthalmic services (for example, cataract removal), musculoskeletal services (for example, arthroscopic joint services), and common gastroenterology procedures (for example, colonoscopy). However, we do not have data on services provided in physician offices. There are 6,236 codes defined as surgical in the CPT system (CPT 10021–69990).

Adding Select Medicine Services

Services such as cardiac catheterization are similar in where they are performed (for example, inpatient hospital setting) and the need for anesthesia to surgical procedures. To identify services in the Medicine category (90000–99999) that should be added to our core services, we focused on codes with RBRVS status codes C, S, or T used in the Medicare outpatient prospective payment system. Status codes are a typology of services. Services with status indicator C are “inpatient only services” that Medicare covers only in an inpatient setting. Status indicators S and T indicate “significant services” (S and T differ on whether a multiple procedure discount applies). Because our goal is to assess the more complex services that will be captured in the databases we are using, we selected an additional 66 codes with these status indicators. These codes are listed in Table A.2.

Adding Select G Codes

As part of the level 2 system in the Healthcare Common Procedure Coding System (CPT), Medicare has a series of G codes (example: G0008) that cover temporary procedures and professional services. Most applicable for this project, G codes have been used by Medicare when a potential CPT code is not applicable for Medicare's purposes. For example, instead of CPT code 45378 (screening colonoscopy), for Medicare submissions screening colonoscopy could be coded as G0121 for lower-risk screening or as G0105 for unremarkable high-risk screening colonoscopy. This is important for our project, as the different databases we use might use different codes for similar services.

We reviewed the list of all G codes and identified those that that had corresponding CPT codes that capture services that were likely to have data in the databases we are using. We list the crosswalk between G codes and CPT codes in Table A.3.

Exclusion Criteria for Codes and Focus on Core Group of Codes

The combination of surgical codes (6,236) and select medical codes (66) gave us 6,302 codes. From the total of 6,302 codes on our list, we then excluded 785 codes that we believe are not applicable to the scope of the project, because they have no work RVUs or it would not be possible to estimate RVUs using our approach. For example, we excluded 510 category II codes that are used for quality measurement. The exclusion criteria and the number of affected codes are in Table A.1. This left us with 5,517 codes.

Of the remaining 5,517 codes, we divided them into three groups: core group (3,179), add-on procedures (194), and exploratory group (2,144). The core group only includes codes with a minimum volume (more than 100 services in a single year within Medicare system) and which we can use our RAND validation method to develop a work RVU. The core group of 3,179 procedures accounts for the vast majority (88.2 percent) of Medicare allowed charges in 2012 across the 6,302 codes. We explain the reasons codes were removed below and in Table A.3.

Dropping Add-On Codes from Core Group

The add-on procedures were categorized into a separate grouping because they should always be performed in conjunction with a primary procedure. In our modeling efforts, our initial objective is to estimate work RVUs for values for services that are performed as the single procedure during a surgical encounter. In Chapter 9 and Appendix F, we explore methods of valuing these add-on codes.

Dropping Low-Volume Codes from Core Group

Of the codes removed, 1,885 were removed because they had less than 100 cases in the Medicare population, including 96 that had no utilization in 2012. We dropped these codes because of concern that we could not capture some of the key characteristics necessary for our models, such as mortality rate and complication rate.

Rationale for Dropping Other Codes from Core Group

We also removed recently deleted codes. However, we crosswalked the key characteristics from these codes to the new or revised codes that replaced the deleted codes so that we were able to value the 2014 codes that were added or split. Some codes are carrier-priced or have restricted coverage, and therefore there is difficulty in determining current CMS payments.

Finally, there are a number of other codes where our validation model structure may not be applicable. We dropped three codes related to intestinal transplant because they have no CCS label and we would have had difficulty putting them into a code grouping. Codes with global period MMM designations and other obstetric procedures are infrequently performed in the Medicare population.

We dropped 25 fluoroscopy codes. These are codes that are often performed in the context of another procedure, and it would have been difficult to value their time independently. Lastly, we dropped 94 codes that were almost always performed outside of a facility with an operating room (< 5 percent of procedures performed in inpatient, ED, ambulatory surgical center, outpatient hospital). The data sources available to us on intra-service time were focused on settings with operating rooms. For procedures performed typically in other settings, the time values in the data sources we use may represent unusual cases and may not be representative of the norm of how the procedure is usually performed.

In the remainder of this report, our main analyses will always focus on the 3,179 core codes. As noted in the final chapter, we can explore applying the RAND validation model to many of the exploratory codes.

Table A.1. Explanation of Procedures Excluded from Core List

Reason for Exclusion	# Removed Codes	# Remaining Codes
Total number of CPT codes (all surgical codes + 66 medicine codes + relevant G codes)	N/A	6,302
Exclusions (785 codes)		
Category II codes (used for quality measurement and not physician payment); end with an F	510	5,792
RBRVS status code B: Bundled code, payment bundled into payment for other services. Separate payment for this provision of these services is never made,	13	5,779
RBRVS status code N: Non-covered service. Code is non-covered under Medicare. If RVUs are shown, they are not used for Medicare payment purposes.	17	5,762
RBRVS status code X: Exclusion by law for Medicare. These items represent an item of service that is not within the definition of physician services for Medicare physician fee schedule payment purposes.	7	5,755
Intestinal transplant codes with no CCS category	3	5,752
No professional component (PC/TC = 3,5)	8	5,744
Intra-service time = 0	159	5,585
Codes deleted 2012–2014	68	5,517
Codes excluded from core list, but what we termed as exploratory codes (2,313 codes)		
RBRVS status code I or E: Code not valid for Medicare purposes. Medicare does not recognize codes assigned this status.	13	5,504
RBRVS status code T: "These services are only paid if there are no other services payable and billed on the same date by the same provider."	1	5,503
RBRVS status code C: "Carrier-priced procedure code." CMS has not established an RVU; typically low-volume services.	27	5,476
RBRVS status code R: Restricted coverage. Special Medicare coverage instructions apply.	64	5,412
No utilization in 2012	96	5,316
Fewer than 100 services in 2012	1,885	3,527
Fluoroscopy code	25	3,502
Procedures typically performed outside facilities where there are operating rooms (POS = <5% in inpatient, ED, ASC, outpatient hospital)	94	3,408
Maternity codes (global = MMM)	11	3,397
Add-on code (global = ZZZ and/or CPT Appendix D)	194	3,203
New code introduced 2012–2014 that could not be crosswalked	24	3,179

Table A.2. Medicine Procedures Added to Procedure List

CPT Code	Short Description
90870	Electroconvulsive therapy
91110	GI tract capsule endoscopy
91111	Esophageal capsule endoscopy
91117	Colon motility 6 hr. study
91120	Rectal sensation test
91122	Anal pressure record
92920	Prq cardiac angioplast 1 art
92921	Prq cardiac angio addl art
92924	Prq card angio/athrect 1 art
92925	Prq card angio/athrect addl
92928	Prq card stent w/angio 1 vsl
92929	Prq card stent w/angio addl
92933	Prq card stent/ath/angio
92934	Prq card stent/ath/angio
92937	Prq revasc byp graft 1 vsl
92938	Prq revasc byp graft addl
92941	Prq card revasc mi 1 vsl
92943	Prq card revasc chronic 1vsl
92944	Prq card revasc chronic addl
92950	Heart/lung resuscitation cpr
92960	Cardioversion electric ext
92961	Cardioversion electric int
92970	Cardioassist internal
92971	Cardioassist external
92973	Prq coronary mech thrombect
92974	Cath place cardio brachytx
92975	Dissolve clot heart vessel
92977	Dissolve clot heart vessel
92986	Revision of aortic valve
92987	Revision of mitral valve
92990	Revision of pulmonary valve
92992	Revision of heart chamber
92993	Revision of heart chamber
92997	Pul art balloon repr percut
92998	Pul art balloon repr percut

CPT Code	Short Description
93451	Right heart cath
93452	Left hrt cath w/ventrclgrphy
93453	R&l hrt cath w/ventriclgrphy
93454	Coronary artery angio s&i
93455	Coronary art/grft angio s&i
93456	R hrt coronary artery angio
93457	R hrt art/grft angio
93458	L hrt artery/ventricle angio
93459	L hrt art/grft angio
93460	R&l hrt art/ventricle angio
93461	R&l hrt art/ventricle angio
93462	L hrt cath trnsptl puncture
93503	Insert/place heart catheter
93505	Biopsy of heart lining
93530	Rt heart cath congenital
93531	R & l heart cath congenital
93532	R & l heart cath congenital
93533	R & l heart cath congenital
93580	Transcath closure of asd
93581	Transcath closure of vsd
93600	Bundle of his recording
93602	Intra-atrial recording
93603	Right ventricular recording
93610	Intra-atrial pacing
93612	Intraventricular pacing
93615	Esophageal recording
93616	Esophageal recording
93618	Heart rhythm pacing
93624	Electrophysiologic study
93642	Electrophysiology evaluation
93660	Tilt table evaluation

Table A.3 Crosswalk of HCPCS G Codes to Relevant CPT Code or Rationale for Dropping

G Code	Description	Crosswalk Code (if any)	Description of Relevant CPT Code (if any) and Rationale
G0104	Ca screen;flexi sigmoidscope	45330	Diagnostic flexible sigmoidoscopy
G0105	Colonoscopy scrn; hi risk ind	45378	Diagnostic colonoscopy
G0106	Colon ca screen; barium enema	Drop	Radiology code
G0120	Colon ca scrn; barium enema	Drop	Radiology code
G0121	Colon ca scrn not hi rsk ind	45378	Diagnostic colonoscopy
G0168	Wound closure by adhesive	Drop	CPT laceration repair codes (12001–12018) restricted to placement/ removal of sutures and staples. Medicare created G0168 because it felt that repairing a laceration with Dermabond was not comparable. Crosswalk not clean.
G0268	Removal of impacted wax md	69210	This code was recently changed to describe ear wax removal with instrumentation. The G code is no longer used.
G0365	Vessel mapping hemo access	93791	The codes could crosswalk to either 93971 (duplex) or single. Duplex would occur only if single does not work.
G0413	Pelvic ring fracture uni/bil	Drop	The G code captures both unilateral and bilateral, while the CPT codes (27216–27218) are unilateral only. In data sources we do not know whether the procedure was unilateral or bilateral and therefore cannot crosswalk cleanly. A low-volume code.
G0414	Pelvic ring fx treat int fix	Drop	See above.
G0415	Open tx post pelvic fxcture	Drop	See above.
G0429	Dermal filler injection(s)	Drop	Cannot crosswalk cleanly to CPT codes 11950–22954, subcutaneous injection or filling material. A low-volume code.

Appendix B. Explanation of Code Groupings

This section provides an explanation of why code groups need to be defined, discusses options for creating the groupings, and summarizes our considerations in choosing how to structure groupings and justification for structure chosen.

Why Do Code Groupings Need to Be Defined?

In this project, RAND's goal was to independently estimate the different components that generate a work RVU for a given service. We recognize that groups of individual services may share characteristics that affect these components of the work and therefore grouping the codes might be useful. We use these code groupings as independent predictors of individual work components, such as pre-service time and post-operative E&M visits. Because code groupings capture services with a similar approach to address similar clinical problems, it could be reasonable to expect that the codes in a group would have similar values for these components. For example, a grouping of neurosurgical codes would likely require a similar amount of time for positioning pre-service.

The code groupings would ideally define a set of services that are clinically similar in both the problem addressed and their complexity. Code groupings based *just* on clinical problems are less amenable to our needs. For example, placement of peripheral intravenous and replacement of indwelling catheter are both services that address a similar problem—putting a catheter or tube into a vein. However, the level of complexity and resources are extremely different, as is the circumstance in which they are performed. One is performed in minutes at the bedside for most patients, and another is often done under sedation, requires monitoring, and is performed in unique circumstances, such as when chemotherapy is needed. While clinically similar procedures are useful in defining groups, we would also like the codes within a code group to have similar levels of complexity and required resources.

We considered three options for defining code groupings:

1. the hierarchical headings that are used to organize the Current Procedural Terminology (CPT) coding system
2. the Clinical Classifications Software (CCS) groupings developed by the Agency for Healthcare Research and Quality to classify CPT services into clinically meaningful procedure categories
3. the Ambulatory Payment Classification (APC) groupings used under the Medicare prospective payment system for hospital outpatient services.

We provide background on each of these grouping systems below. We chose not to include the Berenson-Eggers type of service code (BETOS) system developed by CMS because we felt the categories in the BETOS system were too broad for our purposes.

Three Potential Systems

Current Procedural Terminology (CPT)

The Current Procedural Terminology (CPT) is a nomenclature developed by the American Medical Association to report medical services and services under public and private health insurance programs. The CPT system places each code in a hierarchy of level 1 through level 5 "headings" that are used to organize the system. For example, CPT 11760 (repairing of nail bed) is in "Surgical Procedures on the Nails" (level 3), which is under "Surgical Procedures of the Integumentary System" (level 2), which is under "Surgery" (level 1). Most codes, like CPT 11760, do not have a level 4 or level 5 heading. In this case, the level 3 heading is the most granular code grouping available in CPT. For a minority of CPT codes, there are also level 4 and 5 codes. For example, lacrimal surgery (level 3) is further divided into level 4 subgroups such as "Incision" and "Repair." In creating code groupings we could use any of these levels or some combination.

The CPT system typically divides codes by clinical problem, which is useful. However, there are some limitations. The codes within a heading may require very different levels of resources. Also, the headings are inconsistently provided. As noted above, a minority of codes have level 4 and level 5 headings. Lastly, there are many headings, which create groupings with a small number of codes.

Clinical Classifications Software (CCS)

The CCS is developed by the Agency for Healthcare Research and Quality to provide "a method for classifying Current Procedural Terminology (CPT®) codes and Healthcare Common Procedure Coding System (HCPCS) codes into clinically meaningful procedure categories" (HCUP, 2014). There are 244 categories. Two certified clinical coding specialists assigned codes to CCS categories. Each CPT/HCPCS code was entered into a software program that helps crosswalk CPT codes to ICD-9-CM service codes. Each ICD-9-CM code in the original CCS software was analyzed in order to understand how the ICD-9-CM codes had been assigned to CCS categories. If there were multiple CCS categories for the comparable ICD-9-CM codes, one that best fit the description of the CPT/HCPCS code was chosen. Several principles were used to address concerns. For example, if a code included two services, the one that would result in the greatest morbidity or resource use took precedence. For example, CPT 43305 (esophagoplasty) requires repair of both the esophagus and trachea. Because repair of the esophagus is the more prominent service, it was assigned to the gastrointestinal category. Other principles included assigning codes to a body system associated with the specialty of the physician performing the service. More details on the assignment are provided in the documentation (HCUP, 2014). Final code assignments were reviewed by coding specialists and staff at the Agency for Healthcare Research and Quality.

Like the CPT system, the CCS system typically divides codes by clinical problem, which is useful. It also has a smaller number of groupings, and each code has a single grouping. However, within a grouping, the CCS system includes codes with large differences in complexity and resource use.

Ambulatory Payment Classification (APC)

The outpatient prospective payment system was developed as a method to pay for outpatient hospital care. CMS assigns each outpatient service to one of approximately 800 ambulatory payment classification (APC) groups (CMS, 2012c). The groupings of services (codes) were based on clinical and resource use, and all services within an APC have the same payment rate (MedPac, 2007).

The advantage of the APC system is that the codes capture codes with both similar clinical situations and similar resource use. For example, the placement of a pacemaker and a bronchoscopy should not be grouped together even if resource usage is similar. One major disadvantage is that not all services in our scope of work have an APC designation. Services *never* performed in an outpatient hospital setting lack an APC designation. These are primarily services performed only in inpatient setting. Minor procedures that are packaged into the payment for the primary procedure also lack an APC designation. Also, the APC system captures facility resources in its groupings, while our focus is only on physician work. It is not clear whether this discrepancy is important with respect to surgical procedures. It may be reasonable to expect that differences in the facility resources required for surgical services would also reflect differences in physician intra-service work.

RAND Code Groupings

Our final groupings are a combination of the APC and CCS system. Most of the groupings are based on APCs. We chose the APC as the key backbone for grouping structure because we felt that the APC best captures the clinical coherence we wanted, in that the services were grouped based on both clinical situation and resource use. We felt that facility resource use was a reasonably good proxy for services of similar clinical intensity. For example, we felt that the pre-service time would be most consistent within an APC.

When necessary, we used the CCS system to supplement the APC groupings. For inpatient-only services (Status Indicator C), there are no APC groupings. For those inpatient codes not included in the APC, we chose to use the CCS grouping. For example, two types of radical mastectomy (19305, 19306) are not included in the APC system because Medicare covers them only on an inpatient basis. Because they are in the CCS category “mastectomy,” in our system these two codes are joined together in a new grouping. Other mastectomy procedures that are covered in an outpatient setting are included in a different grouping based on their APC category (Level 2 Breast Surgery).

There are also codes not included in the APC system because they are packaged within the facility fees for the primary procedure under outpatient prospective payment system or are not paid under this system (for example, status indicators E and N). We assigned these codes into separate code groupings based on their CCS category.

Therefore, there are three types of groupings in the overall system.

Table B.1. Three Types of Groupings Used to Classify Procedure Codes in Core Group

Grouping Type	Number of Codes	Percentage of Medicare RVUs	Example of a Grouping Title
From APC	2,508	79.9%	Thrombectomy
Inpatient-only service	688	19.0%	Spinal fusion
Packaged, non-covered service, not performed in outpatient hospital setting	32	1.1%	Suture of skin and subcutaneous tissue

Appendix C. Explanation of Variables Used to Characterize Each Service

In this appendix, we provide more details about how we measured the variables on each procedure that we used in our regression models. We divide up the variables into four groups: (1) time for performance as measured by intra-service time, (2) characteristics of procedure, (3) patient and service complexity, and (4) intensity. A detailed description of intra-service time is provided in Chapter 4. One of the key characteristics of procedure, code grouping, is described in Appendix B. The remaining variables are described below.

Characteristics of Service (Five Variables)

Code Grouping

Description, rationale, and methods are described in Appendix B.

Body System Grouping: CCS Level 1

Brief explanation and rationale: We use body system groupings as predictor variables because this variable might explain some of the variation in positioning or immediate post-operative care.

Method: The Agency for Healthcare Research and Quality produces CCS level 1 and level 2 groupings for ICD-9 procedure codes, but only level 2 groupings for CPT codes (244). The level 1 groupings include 16 different body system groupings of procedures (for example, musculoskeletal procedures). To group the 244 level 2 groupings into 16 level 1 groupings, we used the ICD-9 groupings to generate a crosswalk.

Typical Setting

Brief explanation and rationale: Setting in which services is most often performed. This is captured because it provides some sense of the complexity of the procedure. Incorporating setting is also important for issues such as how much pre-service work is involved.

Method: Among all Medicare beneficiaries who receive a given procedure, what is the most common place of service in which the procedure is performed? The four choices are inpatient, outpatient facility (ambulatory surgical center or outpatient hospital), office and other non-facility settings, or ED.

Global Period

Brief explanation and rationale: Some services are assigned a 10- or 90-day global period in which post-operative visits during that time for that service are bundled into the payment for the surgical procedure. Typically more complex procedures are assigned a global period, and this could be reflected in greater total work and work associated with individual components.

Method: Comes from Addendum B of the 2014 Medicare fee schedule final rule.

Risk Category

Brief explanation and rationale: CMS coverage and payment policies define the settings in which services may be safely provided to Medicare beneficiaries. The continuum of settings include “inpatient only” procedures, services that are payable under the hospital outpatient prospective payment system but not as facility services provided by ambulatory surgical centers, procedures that are payable as ASC facility services, and services that are considered “office-based” procedures for which no ASC facility fee is payable. Typically, the complexity and risks associated with the procedure are reflected in the Medicare policies regarding which settings for which facility fees are payable.

Method: We used Addendum B of the hospital outpatient prospective payment rates to identify which procedures are payable as “inpatient-only” procedures and Addendum BB of the ASC fee schedule to identify procedures that are payable as ASC procedures and those that are payable as “office-based” procedures. Procedures that are covered under the outpatient prospective payment system but not as ASC facility services were identified as “hospital-only” procedures.

Patient and Service Complexity (Eight Variables)

We defined these variables at either the typical place of service or across all places of service. We identified the typical place of service as described in the previous section under “Typical Setting.” For the typical values, we calculated each of the following eight variables only for patients that received the procedure in the typical setting. For the all place of service values, we calculated each of the following eight variables across all instances of a procedure. For some variables—for example, length of stay—we counted 0 days length of stay for procedures performed outside of the inpatient setting when calculating values across places of service. We include more detail of specific steps in our discussion of each variable.

Comorbidity Count

Brief explanation and rationale: We capture the number of age, gender, and comorbidities among patients who receive the procedure in the Medicare population. Together these are important variables for capturing the overall level of illness among the patients who receive the procedure. The assumption is that performing a procedure on a population of patients who are more ill requires more work.

Methods: We identify each time a service was performed in the Medicare population in 2011 using the carrier file to identify when a physician billed for that service. Using diagnosis codes on bills in the carrier file, we identify all patients who in that year (before and after the procedure) had the 22 conditions included in the Charlson index (Charlson et al., 1987). We calculate the median count of Charlson conditions across patients for a given procedure and use that in our regressions.

Age

Brief explanation and rationale: We capture the number of age, gender, and comorbidities among patients who receive the procedure in the Medicare population. Together these are important variables for capturing the overall level of illness among the patients who receive the procedure. The assumption is that performing a procedure on a population of patients who are more ill requires more work.

Methods: Procedures were identified in the carrier file and then linked to the Master Beneficiary Summary File (MBSF) by the patient to obtain the date of birth. Age was calculated as the difference between the date of service and the date of birth. We used the median age of those received the procedure.

Gender

Brief explanation and rationale: We capture the number of age, gender, and comorbidities among patients who receive the procedure in the Medicare population. Gender is a common variable used in risk-adjustment models. Together these are important variables for capturing the overall level of illness among the patients who receive the procedure. The assumption is that performing a procedure on a population of patients who are more ill requires more work.

Methods: We calculate the proportion of female patients among Medicare beneficiaries who receive the service. Procedures were identified in the carrier file and then linked to the MBSF by the patient to obtain the gender.

Length of Stay

Brief explanation and rationale: Length of stay for a hospitalization and how many ICU days after a procedure are markers of the complexity of a procedure. Procedures with a greater length of stay and more ICU days, in theory, should be more work. Length of stay and ICU days are particularly important as a predictor of post-operative visits, in particular those that are inpatient.

Methods: We calculated the length of post-service stay for services provided in an inpatient setting among the Medicare beneficiaries. Services were identified in the carrier file and then linked to the MedPAR file by the patient and date of service to identify a discharge date. The length of stay for a given procedure was calculated as the difference between the discharge date and the date of service. Services performed outside the inpatient setting were given a zero length of stay. We used the median length of stay for a given procedure in our models.

ICU Days

Brief explanation and rationale: Length of stay for a hospitalization and how many ICU days after a procedure are markers of the complexity of a procedure. Procedures with a greater length of stay and more ICU days, in theory, should be more work. Length of stay and ICU days are particularly important as a predictor of post-operative visits, in particular those that are inpatient.

Methods: We calculated the number of days in an ICU for services provided in an inpatient setting among the Medicare beneficiaries. Services were identified in the carrier file and then linked to the MedPAR file by the patient and date of service to identify the number of ICU days. ICU days only during the index hospitalization were consideration. Services performed outside the inpatient setting were given zero ICU days. Median numbers of ICU days were used in the regressions.

Major Complications

Brief explanation and rationale: Major complications and mortality following a procedure were measured as markers of the complexity of a procedure. Procedures that are associated with greater mortality and more complications should, in theory, require greater work.

Methods: Complicating conditions were identified as major complications listed in the Major Complications or Comorbidities (MCC) list maintained for the hospital prospective payment system for inpatient services. The inpatient admission could be on an admission for the initial procedure (if it was performed in the inpatient setting) or any other admission that occurred within 30 days of the procedure. Procedures were identified in the carrier file and then linked to the MedPAR file by the patient and date of service. A complication was defined if there was an inpatient admission with a MCC condition listed anywhere within 30 days of the date of service. If the procedure was performed in the inpatient setting, comorbidities that were present on admission for the initial hospitalization were excluded. Complication rate for a given procedure were used in calculations.

Mortality Rate

Brief explanation and rationale: Major complications and mortality following a procedure were measured as markers of the complexity of a procedure. Procedures that are associated with greater mortality and more complications should, in theory, require greater work.

Methods: We calculated the mortality of Medicare beneficiaries who received the procedure. Procedures were identified in the carrier file and then linked to the MBSF by the patient to obtain the date of death. We calculated the 30-day mortality rate for each procedure by place of service.

Thoracic and Laparoscopic

Brief explanation and rationale: Because these procedures have unique needs in terms of pre-service work, we have created a flag for each type of procedure.

Methods: All procedures on the core list were reviewed by hand, and binary flags were created for both types of procedures.

Intensity (Three Variables)

As described in Chapter 7, we created three variables specific to our models for intra-service intensity and work. The justification of these variables in provided in Chapter 3.

Years of Training

Brief explanation and rationale: In prior theoretical work, one driver of higher intensity was technical skill and training necessary to perform the procedure. To capture this, we calculated for each procedure the years of training among those physicians who performed the procedure.

Methods: For each Medicare-defined specialty, we calculated a minimum years of training. This is defined as the number of years of post–medical school training required for board certification in each physician specialty. Specialties are defined using the same specialty classification in the CMS carrier files. Requirements for board certification were obtained from the American Board of Medical Specialties Guide to Medical Specialties. In some cases we had to refer to the relevant specialty society website. Where the board certification requirements varied, we used the minimum requirement. The final list of specialties with the assigned years of training is provided below in Table C.1. To calculate the average years of training required to carry out a given procedure i, we calculated a weighted average of the years of training for all specialties that provided the service in 2011 (using Medicare data). The weights are the specialty-specific utilization counts.

Malpractice Risk

Brief explanation and rationale: In prior theoretical work, psychological stress was associated with increased intensity of a procedure. One potential driver of psychological stress is malpractice risk.

Methods: We used published malpractice premium risk factors from the proposed rule for the 2015 Medicare Physician Fee Schedule (CMS, 2014a). Table 14 in the proposed rule lists the surgical risk factor for each Medicare specialty category. The risk factors range from 1 (for allergy and immunology) to 13.04 (for neurology and neurosurgery) and are equal to the national average malpractice premium for each specialty divided by the average premium for the specialty with the lowest premiums, allergy and immunology. To construct a malpractice risk index for each procedure, we weighted the specialty risk index by the distribution of specialty volume for all the specialties providing the procedure.

Urgency of Decisionmaking

Brief explanation and rationale: In prior theoretical work, urgency of decisionmaking is a driver of intensity. As a marker of urgency, we measure what fraction of the services performed within the Medicare population occur in an ED or on the first day of a hospital in which the source of admission is emergency.

Methods: We calculated the urgency of procedures received by Medicare beneficiaries. Procedures were identified in the carrier file and then linked to the MedPAR file by the patient and date of service to obtain the source of admission and admission date. Inpatient services were considered urgent if they were admitted from the ED or if the service was performed on the day of an emergent hospital admission. Outpatient services were considered urgent if they were performed in the ED with a subsequent hospital admission.

Table C.1. Medical Specialties and Years of Specialty Training

Code	Specialty	Years of Training
01	General practice	1
02	General surgery	5
03	Allergy and Immunology	5
04	Otolaryngology	5
05	Anesthesiology	4
06	Cardiology	6
07	Dermatology	4
08	Family medicine	3
09	Interventional pain medicine	4
10	Gastroenterology	6
11	Internal medicine	3
12	Osteopathic manipulative therapy	4
13	Neurology	4
14	Neurosurgery	7
16	Obstetrics/gynecology	4
17	Physician/hospice and palliative care	4
18	Ophthalmology	4
19	Oral surgery	5
20	Orthopedic surgery	5
21	Physician/cardiac electrophysiology	7
22	Pathology	3
23	Physician/sports medicine	4
24	Plastic and reconstructive surgery	6
25	Physical medicine and rehabilitation	4
26	Psychiatry	4
27	Physician/geriatric psychiatry	5
28	Colon and rectal surgery	6
29	Pulmonary disease	5
30	Diagnostic radiology	5
33	Thoracic surgery	7
34	Urology	5
36	Nuclear medicine	3
37	Pediatrics	3
38	Geriatrics	4
39	Nephrology	5
40	Hand surgery	6

Code	Specialty	Years of Training
44	Infectious disease	5
46	Endocrinology	5
48	Podiatry	3
66	Rheumatology	5
72	Pain medicine	4
76	Peripheral vascular disease	6
77	Vascular surgery	6
78	Cardiac surgery	6
79	Addiction medicine	3
81	Critical care (intensivist)	5
82	Hematology	5
83	Hematology-oncology	6
84	Preventive medicine	3
85	Maxillofacial surgery	5
86	Neuropsychiatry	5
90	Medical oncology	5
91	Surgical oncology	7
92	Radiation oncology	5
93	Emergency medicine	3
94	Interventional radiology	5
98	Gynecological oncology	7

Appendix D. Estimating Surgical Times from Anesthesia and OR Times

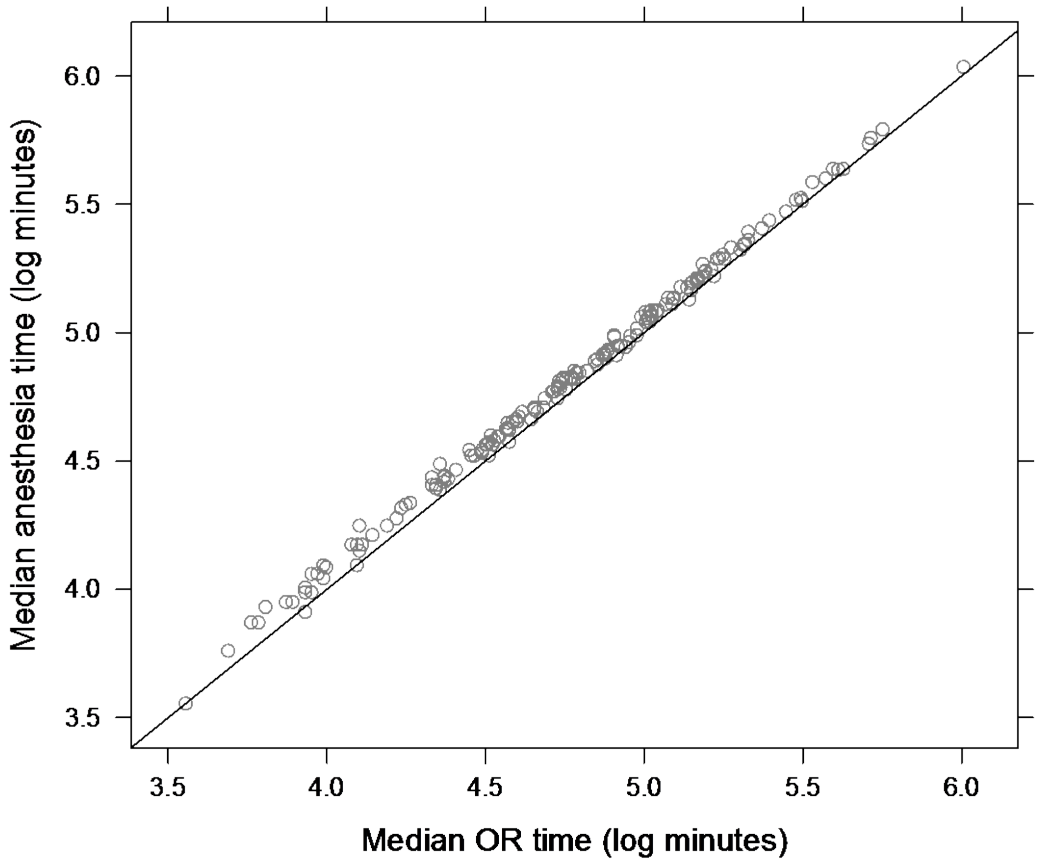
Using the BBM, our goal is to understand the times and intensities that are typically involved with the services that physicians perform. For surgical procedures, there are several data sources that are external to the physician time file that include observed times related to the surgical service. Unfortunately, these data sources do not include the surgical times that are of direct interest for estimating intra-service work. In particular, our available data sources include billed Medicare anesthesia times and OR times. Previous work has shown that anesthesia times can be used to accurately estimate surgical times (as abstracted from patient charts). We extend this work by building models that can be applied to a wider range of surgical procedures and OR times instead of only anesthesia times.

Transforming Anesthesia Times into OR Times

Although we initially expected that anesthesia times and OR times would be essentially interchangeable (on average), our empirical investigations indicate that anesthesia times tend to be somewhat longer than OR times. Our most comprehensive data source (NSAS) for estimating skin-to-skin times from commonly available time elements includes OR times but not anesthesia times. Hence, we begin by estimating a transformation that estimates OR times from anesthesia time information.

We use the National Surgical Quality Improvement Program (NSQIP) to estimate OR times from anesthesia times. NSQIP has relatively sparse coverage of shorter procedures but does contain a variety of time elements. We perform our analysis in the log scale. In a cleaning step, we eliminate any procedures where the OR and anesthesia times differ by more than one unit in the log scale; this removes 422 outlying observations out of the 107,220 available in the analytic sample. Figure D.1 displays median log times in NSQIP for the procedures for which we have at least 30 observations, in both the OR and anesthesia scales. Note that the OR times are typically shorter than the anesthesia times.

Figure D.1. Observed Median Log Times for the Anesthesia and OR Elements in NSQIP, Where at Least 30 Observations Are Available



NOTE: The diagonal line indicates equal anesthesia and OR times.

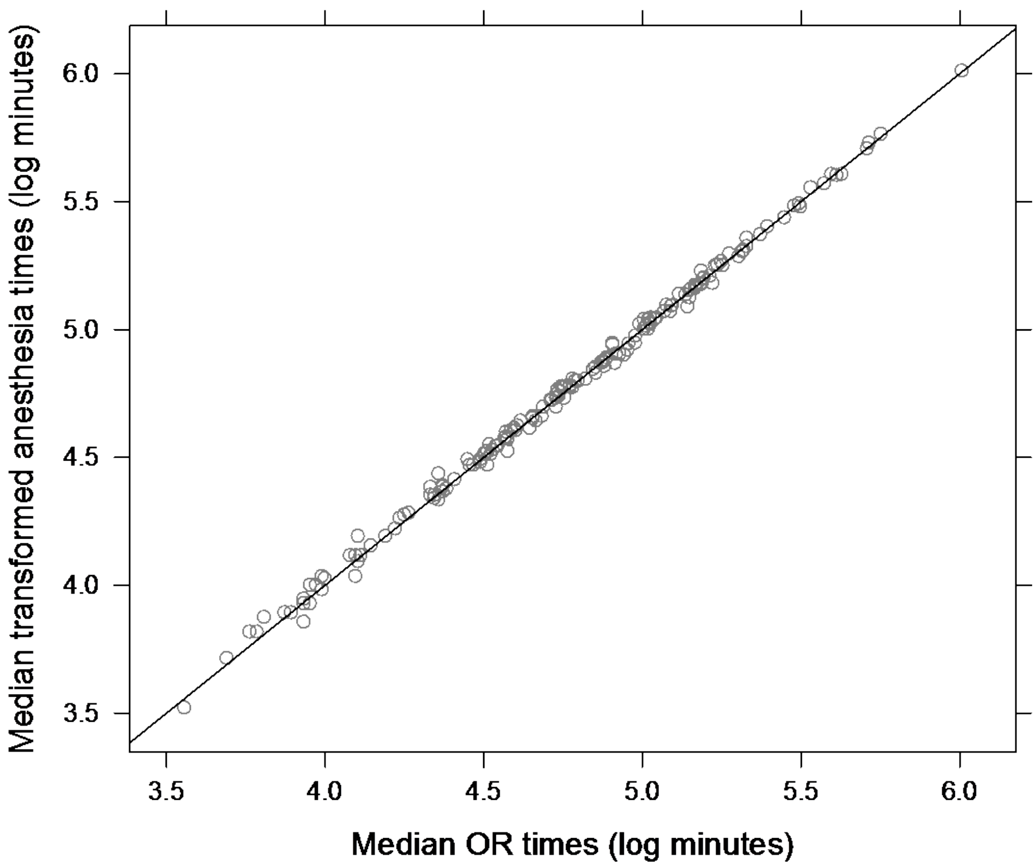
After exploratory analyses, we estimate that anesthesia times and OR times are approximately equal (on average) for procedures with log anesthesia time below 3.0 or above 7.0. In between, OR times tend to be shorter than anesthesia times. Our transformation is given in tabular form below:

Table D.1. Anesthesia and Log Times

Log Anesthesia Time Interval	Estimated Log OR Time
Below 3.0	Log anesthesia time
Between 4.0 and 6.0	1.018 × (log ane time) – 0.131
Above 7.0	Log anesthesia time
3.0 to 4.0 and 6.0 to 7.0	Specified to be continuous and piecewise linear

Although this is a relatively minor adjustment to transform anesthesia times to OR times, without it, over 75 percent of the observed pairs of anesthesia times are longer than the OR times (and are equal for nearly 13 percent). This transformation was estimated using the 2012 NSQIP data. In the 2011 NSQIP data, 53 percent of the transformed anesthesia times are longer than the corresponding OR times; without the transformation, 76 percent of the observed anesthesia times are longer than the OR times. In Figure D.2, note that after applying the transformation, the median log anesthesia times are nearly equal to the observed medians of the log OR times (compare to Figure D.1).

Figure D.2. Median Log Anesthesia Times, After Transformation, Compared to Medians of the Observed OR Times



Transforming OR Times into Surgical Times

Silber et al. (2007, 2011) studied the feasibility of using Medicare anesthesia data to estimate times (anesthesia time, surgical time) that are manually abstracted from patients' charts. They found that anesthesia claim times were highly predictive of the abstracted anesthesia times (5.1 minute median absolute error) and surgical times (13.8 minute median absolute error). Silber's 2011 estimate of the formula to transform from Medicare anesthesia time to surgical time is described by –21.77 min + 0.805 × (anesthesia claim time minutes). The surgical time is a fraction (80.5 percent) of the anesthesia time because anesthesia is given before the surgery begins and concludes after the procedure has finished. This suggests that, in aggregate, the Medicare anesthesia data may be an excellent source of typical service times, even if there are errors for individual clinical encounters.

A key difficulty of applying the Silber transformation directly to our data sources is that it was derived using a relatively narrow set of services. Notably, any observations under 30 minutes were considered "obviously incorrect" for the services under consideration. Indeed, anesthesia times that are 27 minutes or shorter are transformed to *negative* estimated times using the Silber transformation. Since many of our services (particularly in the SPARCS data) are relatively short, we are in need of transformation that is more broadly applicable.

The RAND Transformation: An Extension of the Silber Transformation

A difficulty of the entire project is that we do not have a single data source that covers all of our needs. In order to derive an updated Silber transformation that is appropriate for shorter services, we turn to the 2006 NSAS data. This data source has the key elements of OR time and surgical time, and we build models to estimate surgical times on the basis of OR times. (NSAS only contains ICD-9 codes, and not CPT codes, so we do not entertain it as a data source for our primary time estimates. Also, NSAS has better coverage of shorter surgical services than NSQIP does; otherwise we would have preferred to perform this analysis with those data since it includes OR, anesthesia, and surgical times.) The transformation that we derive most directly applies to OR times rather than anesthesia times; anesthesia times should be transformed to the OR scale before applying the RAND transformation.

We begin our construction of the RAND transformation by performing exploratory analyses. We start with models of the form.

(3.1) estimated surgical time = $\alpha + \beta$ (OR time)

In order to achieve robustness to anomalously long or short observations, we estimate this model using median regression such that "estimated surgical time" means the median surgical time for all observations with a given OR time. In order find an appropriate functional form of our transformation, we estimate model (3.1) on observations whose anesthesia time falls in a 30-minute "window." For example, we estimate it first for observations with OR times that are between 1 and 30 minutes, for observations with OR times between 2 and 31 minutes, and so on. This sliding window approach finds that the estimated slope and intercept terms are equal for all windows that are centered between 15 and 30 minutes. The estimated slopes and intercepts go through a period of transition up to around the 70-minute mark (with the β estimate increasing and the α estimate becoming more negative) after which they become relatively stable. See Figures D.3 and D.4.

Hence, we estimate a transformation that is piecewise linear between 0 and 30 minutes, between 30 and 70 minutes, and greater than 70 minutes. 95 percent of our observations are 132 minutes or shorter. Because we fear that some of the very long recorded times may be misreports (and because such observations would be "high-leverage" points in our regression models), we only estimate the transformation using the lower 95 percent of the observations. We estimate the regression parameters using data in the 0 to 30 minute range and the 70 to 132 minute range and specify the transformation between 30 and 70 minutes so that the transformation is continuous and piecewise linear.

Figure D.3. Estimated Intercept for RAND Transformation, as a Function of the Center of a Sliding 30-Minute Window

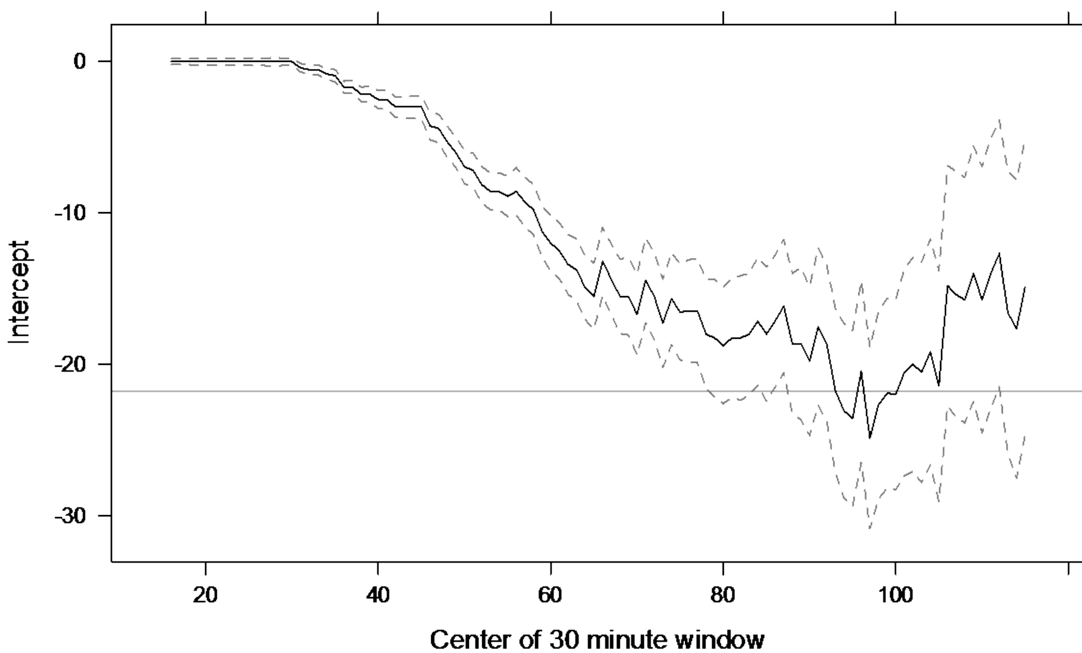
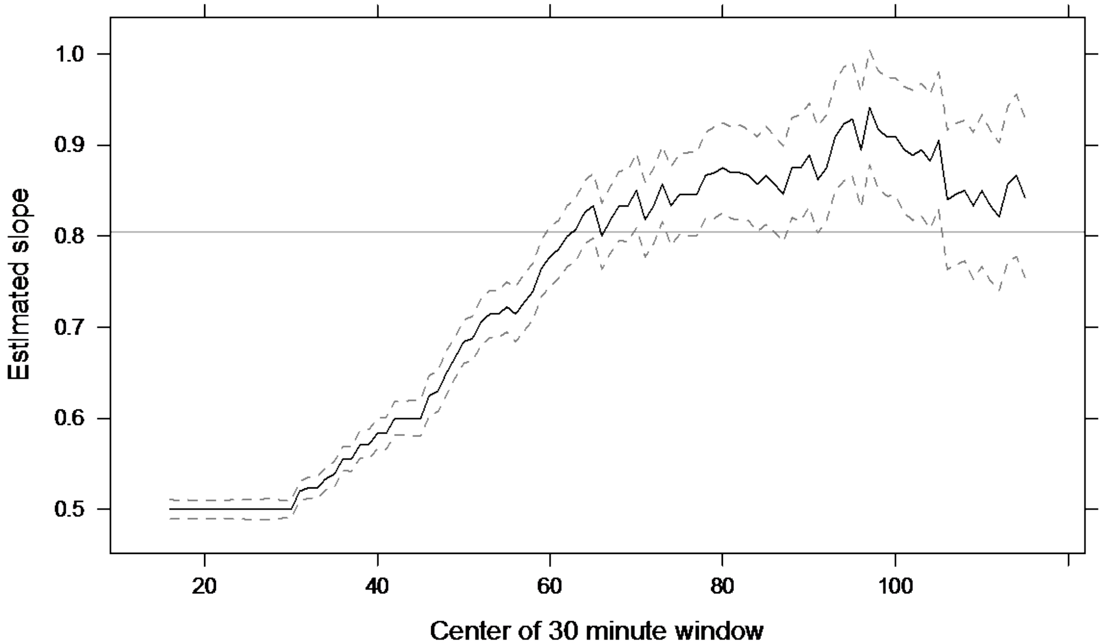


Figure D.4. Estimated Slope for RAND Transformation as a Function of the Center of a Sliding 30-Minute Window



Using all available observations, for the short observations we estimate that an appropriate transformation simply multiplies the OR time by 50 percent. (Because many of the observed surgical times for these shorter observations are exactly half as long as the observed OR time, the median regression estimates are $\alpha = 0$ and $\beta = 0.5$. Such round numbers would be unlikely in standard least squares regression models.) For longer services (70 to 132 minutes), we estimate $\alpha = -17.40$ and $\beta = 0.85$. Note that these are very similar to the Silber transformation, though in all cases our new transformation results in at least slightly longer surgical time estimates. Figure D.5 displays the RAND transformation and the Silber transformation. Even after making the anesthesia to OR transformation, the RAND transformation results in somewhat longer times than the Silber transformation does, though the estimates are rather close, except for short procedures that were not included in the Silber study.

In exploratory analyses to assess the generality of the transformation, we find that the type of anesthesia is estimated to have a small impact on the transformation. Although the CCS level 1 groupings result in more substantial differences, we ultimately do not allow for differences by these body part indicators. We make this decision because a validation exercise using the NSQIP data indicates that our estimates that do not allow for differences by CCS level 1 groupings result in better estimates of the surgical times when applied to anesthesia or OR times.

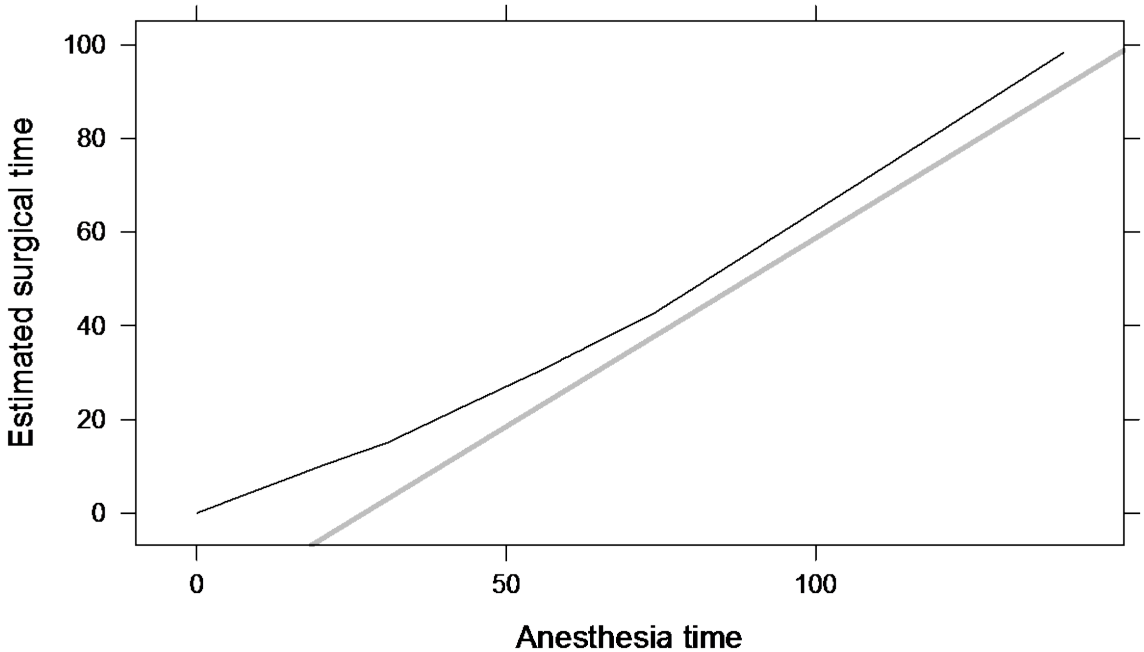
We also examine whether a different transformation was needed depending on what type of anesthesia was administered and whether an anesthesiologist was present. The results in Table

D.2 relate to indicators of various anesthesia options for services with OR times that are between 30 and 132 minutes in length (pooled across CCS groupings) in a flexible (B-spline) model of median surgical time as a function of OR time. Although some indicators are statistically significant, we see that the magnitudes of the estimates are all rather small. (To emphasize, the estimated parameters relate to potential differences in an appropriate transformation by type of anesthesia, not differences in the lengths of the overall OR or surgical time itself.) We deem these differences to be insufficiently large to merit their inclusion in the final RAND transformation.

Table D.2. Estimates of Anesthesia-Related Variables Added as a Constant to a Model Used to Estimate the RAND Transformation for Services with Between 30 and 132 Minutes of OR Time

Type of Anesthesia	Estimate (minutes)	Significant
Topical anesthesia	–0.80	Yes
IV sedation	0.24	No
Monitored anesthesia care	1.56	Yes
Regional epidural	–0.81	Yes
Spinal	–3.93	Yes
Regional retrobular block	0.05	No
General anesthesia	–1.24	Yes
Regional peribulbar block	0.81	Yes
Regional block	–1.79	Yes
With anesthesiologist	1.61	Yes

Figure D.5. RAND Transformation of Billed Anesthesia Times to Surgical Time, as Derived from the 2006 NSAS Data



NOTE: The thin black curve describes the RAND transformation. For comparison, the thick gray line represents the Silber et al. (2011) transformation.

Adjusting for Differences Between SPARCS and Medicare Anesthesia Times

As described above, we argue that it is reasonable to apply the RAND transformation to both anesthesia and SPARCS times after making the anesthesia to OR transformation. After doing so, we would like to pool the two data sources. However, in exploratory analyses we find that there are substantial differences between the two data sources, with SPARCS times typically being shorter. In order to correct for these differences, we subset our analytic files to exclude inpatient records (since SPARCS does not include such observations) and to exclude records where neither general nor regional anesthesia was administered (since the Medicare anesthesia claims are only available when an anesthesiologist submits a bill). For this analytic sample, we estimate a SPARCS indicator for each CCS level 1 grouping in a model of log surgical time (as results from the RAND transformation) with random effects for CPT code. This yields an estimated correction to bring the SPARCS observations into the scale defined as estimated surgical times from Medicare anesthesia bill times. Such an adjustment would be expected to address differences in patient populations, geographic distribution, and time element definitions.

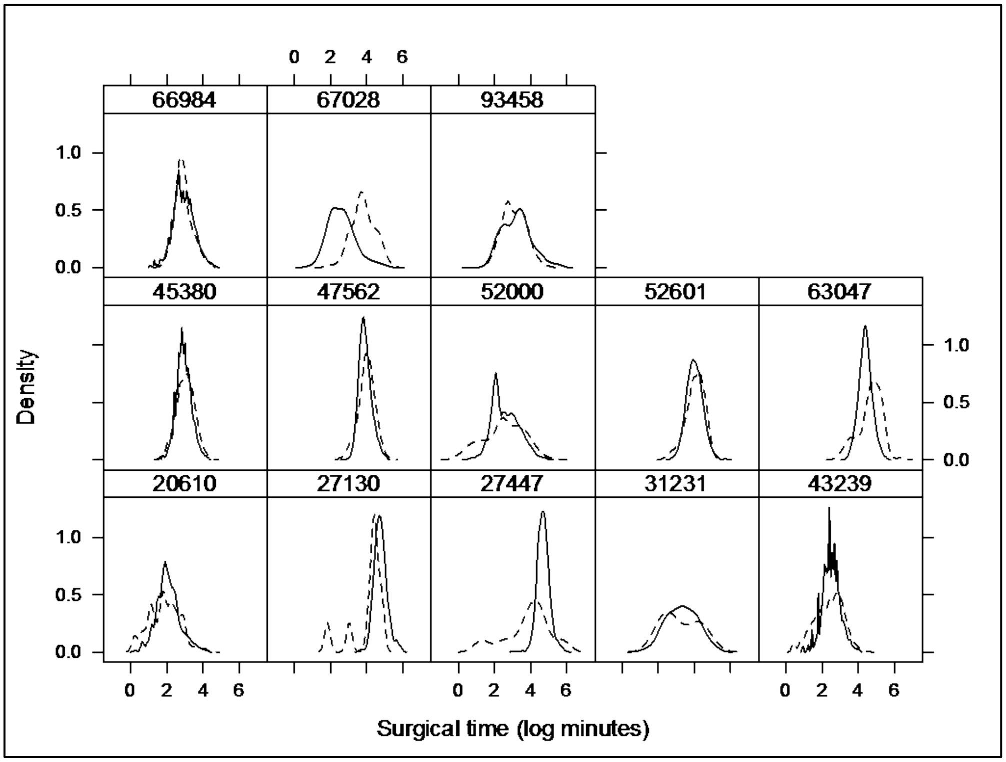
For all CCS level 1 groupings except level 16 (Miscellaneous Services), this adjustment lengthens the SPARCS surgical times. Table D.3 gives these additive adjustments (in the log scale). Figure D.4 displays density histograms of our Top 20 codes after adjustment. Although qualitative differences remain between the two data sources at the CPT level, we believe that the distributions are similar enough to pool.

Ideally, we might like to estimate a SPARCS-to-anesthesia transformation on a procedure-by-procedure basis so that, for instance, the distributions of times for CPT 67028 would be closer. While we have formulated a model that would yield such estimates, it is computationally challenging to estimate. In order to make reasonable corrections when few or no observations are available for a particular procedure in one of the data sources, it is necessary to estimate the corrections for all services in a single model. Given the millions of observations available and thousands of parameters to estimate, estimating this model will likely not be feasible using commercial statistical packages. However, we feel that it will be possible to develop specialized software to produce such estimates in future research.

Table D.3. Estimated Adjustments Added to Log SPARCS OR Time to Bring Such Observations into the Log Medicare Anesthesia Time Scale, by CCS Level 1 Grouping

Body System Grouping (CCS level 1)	Change to Log SPARCS OR Time
Nervous system	0.53
Endocrine system	0.22
Eye	0.37
Ear	0.44
Nose, mouth, and pharynx	0.28
Respiratory system	0.32
Cardiovascular system	0.20
Hemic and lymphatic system	0.24
Digestive system	0.31
Urinary system	0.41
Male genital organs	0.27
Female genital organs	0.56
Obstetrical procedures	0.43
Musculoskeletal system	0.37
Integumentary system	0.29
Miscellaneous services	–0.32

Figure D.6. Density Histograms of Log Surgical Times for Top 20 Services, Where Both Services are Observed in SPARCS (with Anesthesia) and Medicare Anesthesia (Not Inpatient)



NOTE: The solid curves are for Medicare anesthesia and the broken curves are for SPARCS. The SPARCS times have been adjusted to bring them into the scale of Medicare anesthesia surgical time.

Conclusion

In the work described in this appendix, we extend recent research on estimating surgical times from data elements that are more commonly recorded in surgical databases, namely OR times and billed anesthesia time. Our estimated surgical times are slightly longer than those that would result from application of the Silber transformation. After application of the RAND transformation, we find differences between times from the anesthesia and SPARCS files. We therefore adjust the SPARCS times on a CCS level 1 grouping basis, in order to bring the SPARCS estimates into the scale of anesthesia times. Having made this adjustment, we believe that it is reasonable to pool these two data sources, extending our scope beyond procedures that are performed with anesthesia.

Appendix E. A Bayesian Approach to Estimating Intra-Service Times

Introduction

In this project, one of our goals is to estimate the time values for over 3,000 surgical procedures. Although we have millions of time observations in our external data sources, they are distributed very unevenly across the services. For surgical procedures that have few observations in our external data sources, purely empirical estimates (such as sample means) are likely to be very poor estimates of the population values since there is often substantial heterogeneity in times from one surgical encounter to the next. Indeed, we have no external observations for a given procedure in some places of service, making reliance on purely empirical observations impossible. On the other hand, we have hundreds of thousands of time observations for other services. To the extent that these data are clear of data quality issues, the empirical estimates for these services would be expected to be very close to the true population values.

With this in mind, we pursue a method to update times that can produce reasonable estimates when we have no, few, or many external time observations. A mode of statistical inference called Bayesian analysis accommodates just such circumstances. In cases where few external observations are available, existing estimates (e.g., CMS estimates values) are preserved. In cases where much external data is available, the empirical values are adopted. And, when there is an intermediate amount of information available, the reported estimates are a compromise between existing and empirical estimates, smoothly transitioning from low- to high-information settings.

A Brief Overview of Bayesian Inference

Standard statistical inference begins with a statistical model that implicitly or explicitly describes the probability of an outcome of interest, such as the time a surgeon takes to perform a particular surgery, conditional on model parameters. These model parameters describe quantities such as the mean of the distribution of all surgical times for that procedure, for some population of patients and surgeons. For example, our statistical model may be that the observed times are drawn from a normal distribution with an unknown mean and variance.

Bayesian inference starts with the same statistical model (expressed mathematically in what is called a likelihood) and layers on assumptions about the plausibility of model parameters. These assumptions summarize the analyst's beliefs about the model parameters before any data are observed. Because these beliefs about the model parameters are stated before seeing the data, they are called *prior beliefs*. As an example, an experienced surgeon would not need to see any data to be quite confident that CPT 50360 (transplantation of a kidney) requires, on average, more than 15 minutes of skin-to-skin time. Bayesian inference formalizes the use of such prior information in statistical inference, combining prior beliefs and observed data in a logically coherent framework.

Operationally, prior beliefs are quantified as a *prior distribution*. For example, an experienced surgeon's best guess may be that the average skin-to-skin time required to perform the kidney transplant is 180 minutes, but he or she would not be overly surprised if it were as low as 100 minutes or as high as 260 minutes. Note that the prior beliefs relate to the model parameters (such as a mean) rather than individual observations of the outcome of interest (e.g., the time required for a particular surgery). To transform these prior beliefs into a prior distribution, it would be reasonable to approximate them with a normal distribution with mean 180 minutes and standard deviation 40 minutes: Such a prior distribution is centered at the surgeon's best guess and encodes a subjective belief that there is greater than 95 percent probability that the average time is between 100 and 260 minutes.

Having settled on a statistical model (i.e., the likelihood) and having translated prior beliefs into a prior distribution, the estimate of the model parameters is described by Bayes' theorem. Bayes' theorem says that the distribution that describes the updated state of knowledge (after observing the data) is proportional to the product of the likelihood and the prior distribution. This updated distribution is called a *posterior distribution*; the data are used to update the prior distribution into the posterior distribution.

CMS Estimates as Prior Information

Bayesian prior distributions such as the one described above are inherently subjective: One person's prior beliefs may differ substantially from another's. For this reason, applying such assumptions to scientific analyses can, in some fields, be controversial. In physician payment, however, the values reported by CMS are currently in use and therefore trump any individual's personal beliefs. For this reason, we are confident in basing our prior distribution on the current CMS values. In particular, we define the prior distributions in our unadjusted Bayesian models to be normal distributions, centered at the CMS value. Structuring the prior information in this way provides a baseline with which external information can be combined. This process updates CMS time estimates only to the extent that empirical evidence justifies doing so.

Specific Models, Prior Distributions, and Resulting Estimates

As discussed above, for our Bayesian method we need to specify both our statistical model (the likelihood) and the prior distribution. For the statistical model to estimate geometric means, we specify a normal distribution for the log surgical times. (See Appendix D for a discussion of estimating surgical times from OR and Medicare anesthesia billing times.) For the prior distribution, we specify a normal distribution with variance equal to one-thirtieth of the variance of the log time observations around the true mean of the log time values for each CPT code. By specifying the prior in this way, we do not need to estimate the variability of the observations themselves; the Bayesian estimate of the mean log intra-service time is given by a weighted average of the prior estimate (i.e., the current CMS time estimate) and the mean of the log time observations from our external data sources. With the prior distribution's variance defined to be equal to one-thirtieth of the variance of the log times around the true mean, the current CMS time estimate is given the weight of 30 external observations. Given that the RUC historically required 30 survey responses to be considered valid, we believe that this is an appropriate amount of weight to give the current estimates.

Our models that estimate the median times assume that the log times have a Laplace (or double exponential) distribution centered at the median of the log-time distribution. The Laplace distribution is commonly used in Bayesian median regression (Yu and Moyeed, 2001), giving less influence to outlying observations than normal models do. Recent research has shown that specifying this distribution produces valid median estimates even if the true distribution of the outcome variable has another distribution (Sriram et al., 2013). In the case of median regression, we do not have a closed-form expression for the Bayesian estimate like we do for the geometric mean model described above. Instead we use a numerical solver to minimize the function

$$p(\theta) = \frac{(\theta - \theta_0)^2}{30} + \sqrt{2} \sum_{i=1}^{n} | \; t_i - \theta \; |,$$

which is equivalent to maximizing the Bayesian posterior distribution. In this expression, θ is the unknown average of log times whose value we are trying to estimate; θ_0 is the center of the prior distribution, or the current log CMS intra-service time estimate; and, t_i are the log time observations from the external data sources.

Maximizing the posterior distribution in this way yields what are called "maximum *a posteriori*" (MAP) estimates. MAP estimates are popular in situations where the posterior mean is difficult to calculate. This is the case in our situation. Calculating the posterior mean requires integrating over a function that has no closed form integral, and the approximate methods (called Markov Chain Monte Carlo) that are typically used for these models can be very slow to compute, given the sample sizes of external time observations available in this application. And, because we are interested in "best estimates" for intra-service times rather than quantifying the uncertainty in such estimates, little if anything is lost in calculating the MAP estimates.

The Adjusted Bayesian Model

The discussion in the previous section describes the likelihoods, prior distributions, and posterior estimates for what we call the "unadjusted Bayesian model." As motivated in Chapter

4, we are also interested in "adjusted Bayesian models" of intra-service time. By "adjusted," we mean that we adjust the prior distribution so that the prior estimates are not systematically longer or shorter than the CMS values, as may be the case with the unadjusted Bayesian model. Mathematically, we assume that the center of the prior distribution is not at the logarithm of the CMS value, but at the logarithm of some constant k times the CMS value. The constant k is then learned from the data. Technically, $\log(k)$ is itself given a prior distribution, but this is chosen to be noninformative so that no particular value of $\log(k)$ is preferred before seeing the data. We allow for different values of k for each of the three POS classifications used in the intra-service time analyses (Inpatient, Not Inpatient with Anesthesia, and Not Inpatient Without Anesthesia).

The estimated value of k impacts the estimated time for each procedure, and vice versa. Therefore, we take an iterative approach to estimate k and an updated intra-service time for each surgical procedure. As with the unadjusted median model, calculating the full Bayesian solution for this model would be very computationally demanding, due both to the large sample size and the large number of parameters to be estimated. Therefore, we again report Bayesian MAP estimates.

The algorithm to produce the estimates is as follows:

1. Choose a starting value of $\log(k)$.
2. Estimate mean log time values for each service using the sum of $\log(k)$ and the log of the CMS estimate as the center of the prior distribution.
3. Update the estimate of $\log(k)$ to the mean difference between the log time estimates calculated in step 2 and the log of the CMS estimates.
4. Repeat steps 2 and 3 until convergence is achieved.

One can show that the posterior distribution for this model is log-concave, and that each iteration of the updates increases the log posterior density. Through a bounding process, we therefore are able to guarantee that we estimate $\log(k)$ within a tolerance ε of its optimal value. For the reported results, we used a tolerance of $\varepsilon = 0.001$. Given a value of $\log k$, we are then able to estimate the MAP values of the log time estimates exactly, as was the case in the unadjusted Bayesian geometric mean model.

In a situation where the unadjusted prior estimates are systematically long or short, an adjustment such as this one can be quite important. For example, consider a case where—on average—the services are 10 percent shorter than the CMS values, such that the estimated value of k is 90 percent. Say that one service has only one external observation, and that this external observation is exactly 90 percent of the CMS time estimate. In the adjusted model, the prior beliefs and the limited observed data align, making the final estimate 90 percent of the CMS value, as is arguably desired. In the unadjusted model, however, the final estimate for that service would be approximately 99.7 percent of the CMS estimate. Hence this low-volume service would have its time inflated in a relative sense, merely because of its low volume in the external data sources.

Conclusion

In this appendix we describe a method for combining existing time estimates (from the CMS estimates) with external observations. This method is particularly well-suited to the situation where voluminous data are available for some services while few observations are available for others. Whereas some methods might require a certain volume of external data in order to produce new time estimates, the Bayesian approach allows for a smooth transition from the situation where little external data are available (and existing estimates should be given considerable weight) to the situation where extensive external data are available, making the empirical estimates desirable. Further, we develop an "adjusted Bayesian model" where the prior estimate of the typical intra-service time is allowed to deviate systematically from the CMS values. We see that the adjusted Bayesian model maintains relativity between high- and low-volume procedures, where volume is measured in our external data sources.

Appendix F. Multiple Procedures and Add-On Procedures

Chapter 9 discusses results from investigations of appropriate estimates of intra-service work for multiple surgical procedures that are performed in one operative session and add-on procedures. In this appendix we discuss the technical details of the estimation of the models that yield the results discussed in the body of the report.

Methods

We define an individual service's typical service time through its estimated value from observations when it is performed alone. When we turn to pairs of services, we assume that there is a *single multiplier (p)* that can be applied to the secondary service's intra-service time that describes the additional time over the primary service's time.

(1) typical time for pair = typical primary time + p (typical secondary time)

In the current multiple procedure payment reduction system, many secondary codes are reimbursed at half the typical value, which would correspond to p = 50 percent if work replaced time in the above equation. In particular, the surgical services that are included in this policy have CPT modifier 51 equal to 1, 2, or 3, where 3 denotes endoscopic services. For endoscopies, this policy only applies for services that have different "base endoscopies" (CMS, 2012a). For the analyses presented here, we include only pairs of core surgical services that are observed in our Medicare anesthesia analytic file such that both services have CPT modifier 51 equal to 1, 2, or 3. Pairs of endoscopies that have the same base endoscopy are also excluded.

In what follows, "primary" refers to the service that is the longer of the two, as judged by typical estimates from times when they are performed alone. The primary/secondary classification does not necessarily reflect clinical importance or work RVUs.

Qualitatively, at one extreme, setting $p = 0$ implies that secondary services add no time relative to the typical time required for the primary service; if p = 100 percent, then the typical time for the pair of services simply equals the sum of the typical times, implying no efficiency of scale.

A method for estimating p that is compatible with the geometric means used to produce our preferred time values chooses the value of p that

(2) log(typical time for pair) ≈ log(typical primary time + p (typical secondary time)).

In order to best estimate p in this formulation, we reframe (2) as

(3) log(time for pair) = log(typical primary time + p (typical secondary time)) + ε,

where ε is a normally distributed, mean zero error. As in linear regression, we can estimate p by minimizing the sum of squared differences between the log observed time and the log of the time estimate implied by any value of p. Although this is a nonlinear regression model, it is straightforward to estimate its single parameter, and the nonlinear least squares function "nls" in the R statistical computing environment is able to estimate confidence intervals for this multiplier.

Multiple Service Adjustments via Medians

In the discussions above, the multiplier p is estimated through equation (3): We find the p such that the log time of the joint service is well-approximated by the log of the sum of the two individual services' estimated times, where the secondary service's time is discounted by the factor p. An alternative approach is to consider the time estimates to be median intra-service time estimates. Using the definition of a median, we expect approximately half of the observed times to be above the estimate, and half below. In this way, we can select a multiplier p that achieves this split. A benefit of this approach is that outlying observations do not have overly strong influence on the estimates of p.

Add-On Estimates

Our methodology for estimating intra-service times for add-on services is very similar to our main method for estimating the multiple procedure correction, as described above. In this case, we want to estimate the additional time that is required to perform a base procedure and an add-on procedure together, relative to the base procedure alone. We perform this least squares estimation separately for all add-ons for which we have at least 30 observations with one or more of our core codes. The differences between this method and the least squares method described above are that (1) the estimation is performed separately for each add-on code, rather than estimating a universal multiplier, and (2) there is a Bayesian component so that the existing CMS estimate of the time for each add-on receives the weight of 30 observations. The estimation itself proceeds by optimizing the quantity:

$$\sum(\log(y_j) - (\log(t_{i(j)} + t_{addon}))^2 + 30(\log(t_{addon}) - (\log(t_{cms}))^2$$

In this quantity, y_i is the observed intra-service time for the jth observation that contains a particular add-on procedure, $t_{i(j)}$ is the all POS RAND time for the base procedure for observation j, t_{addon} is the time of the add-on procedure whose value is being estimated, and t_{cms} is the existing CMS time file value for the add-on code being considered. This quantity is non-linear in t_{addon} , so we optimize numerically.

Appendix G. Regression Output from Pre-Service, Intra-Service, Post-Service, and Total Work Prediction Models

Appendix G reports estimated coefficients for the regression models described in Chapters 5 through 8. Tables G.1.A and G.1.B report estimated coefficients for the pre-service and immediate post-service regressions described in Chapter 5 for the typical place of service (Table G.1.A) and across places of service (Table G.1.B). Table G.2 reports estimated coefficients for the post-operative E&M regressions described in Chapter 6. Table G.3 reports estimated coefficients for intra-service work regressions and intra-service intensity regressions described in Chapter 7, as well as estimated coefficients for the total work regressions described in Chapter 8.

Table G.1.A. Pre-Service and Post-Service Regression Coefficients, Typical POS

Variable	Measure	Pre-Service Evaluation	Pre-Service Positioning	Pre-Service Scrub	Immediate Post-Service
Log RAND intra-service time estimate (mins.)	Estimate	0.441***	0.226***	1.108***	0.430***
Log RAND intra-service time estimate (mins.)	Standard error	(0.0239)	(0.0626)	(0.201)	(0.0308)
Risk category (omitted ASC office)					
ASC OPPS	Estimate	0.210***	0.268***	1.715***	0.207***
ASC OPPS	Standard error	(0.0377)	(0.0974)	(0.335)	(0.0339)
Hospital only	Estimate	0.329***	0.0894	1.483***	0.263***
Hospital only	Standard error	(0.0590)	(0.144)	(0.506)	(0.0762)
Inpatient only	Estimate	0.355***	0.507***	2.634***	0.267***
Inpatient only	Standard error	(0.0638)	(0.157)	(0.501)	(0.0584)
CCS body system (omitted nervous system.)					
Endocrine system	Estimate	–0.0150	0.0969	0.763	0.446***
Endocrine system	Standard error	(0.0911)	(0.141)	(0.862)	(0.142)
Eye	Estimate	–0.190**	–0.394	–0.845	–0.0674
Eye	Standard error	(0.0756)	(0.290)	(0.590)	(0.0866)
Ear	Estimate	–0.445***	–0.626	–0.0895	–0.281***
Ear	Standard error	(0.111)	(0.435)	(0.647)	(0.108)
Nose/mouth	Estimate	–0.249***	–0.0810	–0.398	0.0140
Nose/mouth	Standard error	(0.0843)	(0.212)	(0.655)	(0.0853)
Respiratory system	Estimate	–0.157**	0.201	0.771	0.113

Variable	Measure	Pre-Service Evaluation	Pre-Service Positioning	Pre-Service Scrub	Immediate Post-Service
Respiratory system	Standard error	(0.0729)	(0.217)	(0.657)	(0.0861)
Cardiovascular	Estimate	0.0658	–0.301*	–1.510***	0.192*
Cardiovascular	Standard error	(0.0722)	(0.162)	(0.573)	(0.0981)
Hemic/lymphatic	Estimate	0.276**	0.193	–0.445	0.122
Hemic/lymphatic	Standard error	(0.110)	(0.134)	(0.943)	(0.112)
Digestive system	Estimate	0.161**	0.0102	–0.615	0.146*
Digestive system	Standard error	(0.0645)	(0.134)	(0.629)	(0.0792)
Urinary system	Estimate	0.0160	0.134	0.815	0.0145
Urinary system	Standard error	(0.0729)	(0.188)	(0.741)	(0.0946)
Male genital organs	Estimate	0.0987	0.168	0.427	0.192*
Male genital organs	Standard error	(0.0856)	(0.253)	(0.680)	(0.0988)
Female genital organs	Estimate	0.0366	–0.146	–1.947*	0.0683
Female genital organs	Standard error	(0.0834)	(0.178)	(1.014)	(0.0983)
Musculoskeletal	Estimate	–0.0742	0.0908	0.460	0.0470
Musculoskeletal	Standard error	(0.0645)	(0.140)	(0.570)	(0.0925)
Integumentary	Estimate	–0.189**	–0.238*	–1.220**	–0.0844
Integumentary	Standard error	(0.0743)	(0.143)	(0.603)	(0.0858)
Miscellaneous	Estimate	–0.135	–0.225	0.162	–0.0591
Miscellaneous	Standard error	(0.100)	(0.217)	(0.691)	(0.0954)
Global period (omitted 000)					
010	Estimate	–0.0136	0.0525	0.164	–0.0850**
010	Standard error	(0.0498)	(0.167)	(0.433)	(0.0418)
090	Estimate	0.127***	0.346***	1.906***	0.0386
090	Standard error	(0.0425)	(0.106)	(0.330)	(0.0461)
XXX	Estimate	–0.137	0.138	0.683	–0.0573
XXX	Standard error	(0.226)	(0.193)	(0.538)	(0.200)
Median LOS	Estimate	–0.00606			–0.0229***
Median LOS	Standard error	(0.00727)			(0.00808)
Median ICU days	Estimate	–0.000745			0.00846
Median ICU days	Standard error	(0.00991)			(0.0129)
Percentage female	Estimate	0.188***			0.168***
Percentage female	Standard error	(0.0633)			(0.0632)
Median patient age	Estimate	0.000908			–0.00189
Median patient age	Standard error	(0.00266)			(0.00214)
Median comorbidity count	Estimate	0.0233			0.0219
Median comorbidity count	Standard error	(0.0160)			(0.0158)
Complication rate	Estimate	0.307			0.999**

Variable	Measure	Pre-Service Evaluation	Pre-Service Positioning	Pre-Service Scrub	Immediate Post-Service
Complication rate	Standard error	(0.379)			(0.438)
Mortality rate	Estimate	–0.726			–0.738
Mortality rate	Standard error	(0.478)			(0.478)
Urgent rate	Estimate	–0.0840			–0.137
Urgent rate	Standard error	(0.0839)			(0.0844)
Laparoscopic flag	Estimate	–0.0655	0.382***	–2.736***	–0.182***
Laparoscopic flag	Standard error	(0.0543)	(0.0909)	(0.447)	(0.0529)
Thoracic flag	Estimate	–0.110	0.241***	0.328	0.0185
Thoracic flag	Standard error	(0.0935)	(0.0765)	(0.278)	(0.156)
Constant	Estimate	1.194***	0.736***		1.153***
Constant	Standard error	(0.213)	(0.249)		(0.189)
Ordered logit cut 1	Estimate			4.419***	
Ordered logit cut 1	Standard error			(0.854)	
Ordered logit cut 2	Estimate			5.471***	
Ordered logit cut 2	Standard error			(0.870)	
Ordered logit cut 3	Estimate			8.419***	
Ordered logit cut 3	Standard error			(0.908)	
Ordered logit cut 4	Estimate			9.105***	
Ordered logit cut 4	Standard error			(0.897)	
Random Effect var(_cons)	Estimate	0.0517***	0.179***	1.657***	0.0496***
Random Effect var(_cons)	Standard error	(0.00850)	(0.0455)	(0.499)	(0.00876)
Observations		3,141	1,171	1,942	3,150
Number of groups		299	221	245	298

NOTE: Robust standard errors clustered at the RAND code group level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table G.1.B. Pre-Service and Post-Service Regression Coefficients, Across All POS

Dependent Variable	Measure	Pre-Service Evaluation	Pre-Service Positioning	Pre-Service Scrub	Immediate Post-Service
Log RAND intra-service time estimate (mins)	Estimate	0.453***	0.225***	1.076***	0.446***
Log RAND intra-service time estimate (mins)	Standard error	(0.0244)	(0.0630)	(0.206)	(0.0324)
Risk category (omitted ASC office)					
ASC OPPS	Estimate	0.210***	0.271***	1.750***	0.197***
ASC OPPS	Standard error	(0.0375)	(0.0966)	(0.330)	(0.0328)
Hospital only	Estimate	0.335***	0.105	1.582***	0.265***
Hospital only	Standard error	(0.0600)	(0.142)	(0.499)	(0.0729)
Inpatient only	Estimate	0.341***	0.519***	2.724***	0.255***
Inpatient only	Standard error	(0.0664)	(0.155)	(0.492)	(0.0574)
CCS (omitted nervous sys.)					
Endocrine system	Estimate	–0.0319	0.0889	0.749	0.429***
Endocrine system	Standard error	(0.0928)	(0.145)	(0.871)	(0.141)
Eye	Estimate	–0.185**	–0.399	–0.843	–0.0566
Eye	Standard error	(0.0760)	(0.293)	(0.585)	(0.0837)
Ear	Estimate	–0.441***	–0.627	–0.110	–0.272**
Ear	Standard error	(0.113)	(0.436)	(0.637)	(0.106)
Nose/mouth	Estimate	–0.256***	–0.0900	–0.411	0.0136
Nose/mouth	Standard error	(0.0836)	(0.214)	(0.651)	(0.0847)
Respiratory system	Estimate	–0.181**	0.198	0.770	0.0823
Respiratory system	Standard error	(0.0747)	(0.218)	(0.655)	(0.0850)
Cardiovascular	Estimate	0.0539	–0.306*	–1.504***	0.176*
Cardiovascular	Standard error	(0.0734)	(0.165)	(0.572)	(0.0983)
Hemic/lymphatic	Estimate	0.283***	0.191	–0.439	0.126
Hemic/lymphatic	Standard error	(0.110)	(0.140)	(0.929)	(0.112)
Digestive system	Estimate	0.156**	–0.00110	–0.641	0.132*
Digestive system	Standard error	(0.0663)	(0.136)	(0.623)	(0.0788)
Urinary system	Estimate	0.00572	0.133	0.836	0.0150
Urinary system	Standard error	(0.0704)	(0.189)	(0.732)	(0.0953)
Male genital organs	Estimate	0.105	0.166	0.447	0.197**
Male genital organs	Standard error	(0.0871)	(0.254)	(0.678)	(0.0989)
Female genital organs	Estimate	0.0301	–0.151	–1.907*	0.0586
Female genital organs	Standard error	(0.0849)	(0.178)	(1.005)	(0.0974)
Musculoskeletal	Estimate	–0.0789	0.0840	0.457	0.0358

Dependent Variable	Measure	Pre-Service Evaluation	Pre-Service Positioning	Pre-Service Scrub	Immediate Post-Service
Musculoskeletal	Standard error	(0.0647)	(0.141)	(0.568)	(0.0914)
Integumentary	Estimate	–0.188**	–0.234	–1.189**	–0.0837
Integumentary	Standard error	(0.0731)	(0.144)	(0.598)	(0.0844)
Miscellaneous	Estimate	–0.126	–0.227	0.177	–0.0570
Miscellaneous	Standard error	(0.100)	(0.225)	(0.698)	(0.0954)
Global period (0-day omitted)					
010	Estimate	–0.0174	0.0433	0.131	–0.0871**
010	Standard error	(0.0497)	(0.167)	(0.433)	(0.0415)
090	Estimate	0.116***	0.343***	1.899***	0.0352
090	Standard error	(0.0424)	(0.105)	(0.330)	(0.0466)
XXX	Estimate	–0.147	0.151	0.712	–0.0714
XXX	Standard error	(0.218)	(0.181)	(0.524)	(0.185)
Median LOS	Estimate	–0.00551			–0.0237***
Median LOS	Standard error	(0.00820)			(0.00865)
Median ICU days	Estimate	0.00215			0.0122
Median ICU days	Standard error	(0.0120)			(0.0191)
Percentage female	Estimate	0.203***			0.192***
Percentage female	Standard error	(0.0659)			(0.0658)
Median patient age	Estimate	0.00124			–0.00138
Median patient age	Standard error	(0.00265)			(0.00222)
Median comorbidity count	Estimate	0.0227			0.0136
Median comorbidity count	Standard error	(0.0160)			(0.0184)
Complication rate	Estimate	0.472			1.154**
Complication rate	Standard error	(0.385)			(0.481)
Mortality rate	Estimate	–0.965*			–0.861
Mortality rate	Standard error	(0.494)			(0.542)
Urgent rate	Estimate	–0.151			–0.218*
Urgent rate	Standard error	(0.137)			(0.129)
Laparoscopic flag	Estimate	–0.0629	0.385***	–2.718***	–0.182***
Laparoscopic flag	Standard error	(0.0532)	(0.0914)	(0.445)	(0.0548)
Thoracic flag	Estimate	–0.134	0.236***	0.318	0.00200
Thoracic flag	Standard error	(0.0955)	(0.0778)	(0.283)	(0.159)
Constant	Estimate	1.120***	0.740***		1.051***
Constant	Standard error	(0.212)	(0.255)		(0.197)
Ordered logit cut 1	Estimate			4.345***	
Ordered logit cut 1	Standard error			(0.866)	

Dependent Variable	Measure	Pre-Service Evaluation	Pre-Service Positioning	Pre-Service Scrub	Immediate Post-Service
Ordered logit cut 2	Estimate			5.393***	
Ordered logit cut 2	Standard error			(0.884)	
Ordered logit cut 3	Estimate			8.337***	
Ordered logit cut 3	Standard error			(0.921)	
Ordered logit cut 4	Estimate			9.020***	
Ordered logit cut 4	Standard error			(0.911)	
Random Effect var(_cons)	Estimate	0.0519***	0.181***	1.629***	0.0476***
Random Effect var(_cons)	Standard error	(0.00864)	(0.0464)	(0.493)	(0.00834)
Observations		3,141	1,171	1,942	3,150
Number of groups		299	221	245	298

NOTE: Robust standard errors are clustered at the RAND code group level in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table G.2. Post-Operative E&M Regression Coefficients, Typical POS

Variables	Measure	Outpatient	Inpatient	Critical Care	Discharge
Log RAND intra-service time estimate (mins.)	Estimate	0.356***	0.586***	0.598**	1.236***
	Standard error	(0.0344)	(0.0617)	(0.284)	(0.206)
Risk category (omitted ASC office)					
ASC OPPS	Estimate	0.0899*		–1.671	
	Standard error	(0.0476)		(1.560)	
Hospital only	Estimate	0.0569	0.0557	12.37***	
	Standard error	(0.0705)	(0.127)	(1.297)	
Inpatient only	Estimate	–0.0598	0.671***	11.48***	
	Standard error	(0.0727)	(0.115)	(1.309)	
CCS body system (omitted nervous system)					
Endocrine system	Estimate	0.0139	–0.443**		
	Standard error	(0.115)	(0.221)		
Eye	Estimate	0.789***			
	Standard error	(0.0996)			
Ear	Estimate	0.335***	–0.333**		
	Standard error	(0.0850)	(0.136)		
Nose/mouth	Estimate	0.331***	0.0457		
	Standard error	(0.115)	(0.297)		
Respiratory system	Estimate	0.0311	0.182		
	Standard error	(0.198)	(0.149)		
Cardiovascular	Estimate	–0.503***	–0.318**		
	Standard error	(0.103)	(0.159)		
Hemic/lymphatic	Estimate	0.145	–0.234		
	Standard error	(0.108)	(0.192)		
Digestive system	Estimate	–0.0599	–0.122		

	Standard error	(0.0851)	(0.140)		
Urinary system	Estimate	0.111	–0.276*		
	Standard error	(0.0970)	(0.167)		
Male genital organs	Estimate	0.0579	–0.555***		
	Standard error	(0.104)	(0.208)		
Female genital organs	Estimate	–0.212**	–0.340**		
	Standard error	(0.0958)	(0.167)		
Musculoskeletal	Estimate	0.644***	-0.255*		
	Standard error	(0.0703)	(0.143)		
Integumentary	Estimate	0.445***	–0.200		
	Standard error	(0.0921)	(0.165)		
Miscellaneous	Estimate	0.940*	–0.372*		
	Standard error	(0.538)	(0.210)		
Global: 090 (omitted 10)	Estimate	2.428***	1.078***		
	Standard error	(0.191)	(0.397)		
Median LOS	Estimate	0.00821	0.108***	0.00559	0.773***
	Standard error	(0.00881)	(0.0126)	(0.0708)	(0.167)
Median ICU days	Estimate	–0.0191	–0.0758***	–0.0443	–0.991***
	Standard error	(0.0171)	(0.0205)	(0.0932)	(0.305)
Percentage female	Estimate	–0.0516	–0.00657	–1.497	–1.113***
	Standard error	(0.0730)	(0.154)	(1.022)	(0.418)
Median patient age	Estimate	0.000494	–0.00879	–0.0897***	–0.0783***
	Standard error	(0.00264)	(0.00554)	(0.0246)	(0.0283)
Median comorbidity count	Estimate	0.00868	0.0332	–0.174	0.216*
	Standard error	(0.0174)	(0.0262)	(0.274)	(0.118)
Complication rate	Estimate	–0.222	0.658	3.382	–9.231
	Standard error	(0.389)	(0.443)	(2.545)	(6.968)
Mortality rate	Estimate	0.245	–0.135	3.751**	17.68**
	Standard error	(0.455)	(0.511)	(1.841)	(8.993)
Urgent rate	Estimate	0.0950	1.085***	–1.274	–1.730*
	Standard error	(0.0796)	(0.287)	(0.865)	(1.015)
Constant	Estimate	–2.861***	–2.934***	–8.310***	2.301
Constant	Standard error	(0.250)	(0.648)	(1.974)	(2.045)

Observations	2,467	962	170	1,855
Number of groups	236	149		214

NOTE: Robust standard errors are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table G.3. Coefficients from Prediction Models for Intra-Service Work, Intra-Service Intensity, and Total Work

Variable	Measure	Intra-Service Work, Typical POS	Intra-Service Work, All POS	Intra-Service Intensity, CMS Time File	Total Work
Log RAND intra-service time estimate (mins)	Estimate	0.946***	0.977***	0.0190	0.685***
Log RAND intra-service time estimate (mins)	Standard error	(0.0457)	(0.0444)	(0.0335)	(0.0275)
Risk category (omitted ASC office)					
ASC OPPS	Estimate	0.132**	0.120**	0.0985**	0.130***
	Standard error	(0.0585)	(0.0551)	(0.0418)	(0.0341)
Hospital only	Estimate	0.346***	0.293***	0.297***	0.291***
	Standard error	(0.0803)	(0.0797)	(0.0710)	(0.0476)
Inpatient only	Estimate	0.332***	0.243***	0.316***	0.400***
	Standard error	(0.0920)	(0.0931)	(0.0753)	(0.0609)
CCS (omitted nervous sys.)					
Endocrine system	Estimate	0.0776	–0.00281	–0.0981	0.0135
	Standard error	(0.123)	(0.110)	(0.116)	(0.0904)
Eye	Estimate	0.226*	0.241*	0.347***	0.0988
Eye	Standard error	(0.126)	(0.136)	(0.0982)	(0.0992)
Ear	Estimate	0.266**	0.278**	0.372***	0.00941
Ear	Standard error	(0.121)	(0.121)	(0.103)	(0.0791)
Nose/mouth	Estimate	–0.0949	–0.123	–0.0825	–0.117*
Nose/mouth	Standard error	(0.0889)	(0.0922)	(0.0952)	(0.0665)
Respiratory system	Estimate	–0.0555	–0.141	–0.0754	–0.115
Respiratory system	Standard error	(0.102)	(0.110)	(0.115)	(0.0721)
Cardiovascular	Estimate	0.0762	0.0967	0.0807	–0.0308
Cardiovascular	Standard error	(0.112)	(0.116)	(0.0990)	(0.0847)
Hemic/lymphatic	Estimate	–0.108	–0.0710	–0.0850	–0.0742
Hemic/lymphatic	Standard error	(0.130)	(0.128)	(0.139)	(0.0936)
Digestive system	Estimate	0.0147	0.0195	–0.0505	0.0258
Digestive system	Standard error	(0.0921)	(0.0969)	(0.0842)	(0.0647)
Urinary system	Estimate	0.0521	0.0352	0.0443	0.00268
Urinary system	Standard error	(0.0896)	(0.0947)	(0.0823)	(0.0748)
Male genital organs	Estimate	0.0312	0.0612	0.0787	–0.00656
Male genital organs	Standard error	(0.116)	(0.119)	(0.0995)	(0.0748)
Female genital organs	Estimate	–0.0303	–0.0794	0.0268	–0.0347
Female genital organs	Standard error	(0.0940)	(0.0979)	(0.0948)	(0.0694)
Musculoskeletal	Estimate	–0.0519	–0.0833	–0.113	–0.0431
Musculoskeletal	Standard error	(0.0871)	(0.0930)	(0.0878)	(0.0657)
Integumentary	Estimate	–0.100	–0.0888	–0.120	–0.100
Integumentary	Standard error	(0.120)	(0.110)	(0.0884)	(0.0817)
Miscellaneous	Estimate	–0.0906	–0.0765	–0.160	–0.151

Miscellaneous	Standard error	(0.143)	(0.142)	(0.124)	(0.0941)
Global period (0-day omitted)					
010	Estimate	–0.242***	–0.246***	–0.284***	0.188***
010	Standard error	(0.0576)	(0.0542)	(0.0439)	(0.0363)
090	Estimate	0.0175	–0.0392	–0.215***	0.591***
090	Standard error	(0.0646)	(0.0610)	(0.0419)	(0.0444)
XXX	Estimate	–0.0242	–0.0539	–0.103	–0.251*
XXX	Standard error	(0.160)	(0.147)	(0.116)	(0.138)
Median LOS	Estimate	–0.0280***	–0.00729	–0.0220***	–0.00371
Median LOS	Standard error	(0.0101)	(0.0123)	(0.00853)	(0.00588)
Median ICU days	Estimate	0.0403***	0.0471***	0.0283***	0.0168*
Median ICU days	Standard error	(0.0119)	(0.0149)	(0.00912)	(0.00935)
Percentage female	Estimate	0.0710	0.125	0.0754	0.0584
Percentage female	Standard error	(0.0898)	(0.0833)	(0.0539)	(0.0570)
Median patient age	Estimate	0.00511	0.00478	0.00553***	0.00257
Median patient age	Standard error	(0.00343)	(0.00326)	(0.00205)	(0.00224)
Median comorbidity count	Estimate	–0.509	–0.896*	0.238	–0.00820
Median comorbidity count	Standard error	(0.427)	(0.535)	(0.400)	(0.254)
Complication rate	Estimate	0.0256	0.0275	–0.00603	0.0327***
Complication rate	Standard error	(0.0200)	(0.0189)	(0.0177)	(0.0121)
Mortality rate	Estimate	0.561	0.482	0.0117	0.00734
Mortality rate	Standard error	(0.423)	(0.423)	(0.306)	(0.221)
Urgent rate	Estimate	0.0629	0.216	0.0426	0.157*
Urgent rate	Standard error	(0.157)	(0.241)	(0.180)	(0.0868)
Malpractice risk	Estimate	–0.0118	–0.00713	0.00821	0.00912
Malpractice risk	Standard error	(0.0173)	(0.0167)	(0.0126)	(0.0116)
Training index	Estimate	0.0207	0.00584	–0.0118	0.0131
Training index	Standard error	(0.0150)	(0.0152)	(0.0123)	(0.0102)
ED percentage	Estimate	0.00452	–0.113	–0.0137	–0.266*
ED percentage	Standard error	(0.259)	(0.218)	(0.176)	(0.151)
Constant	Estimate	–3.015***	–3.032***	–3.223***	–1.683***
Constant	Standard error	(0.354)	(0.336)	(0.217)	(0.230)
Observations		3,179	3,179	3,179	3,179
Number of groups		303	303	303	303

Appendix H. Comparison of Unweighted Means in RAND Estimates to CMS Estimates

Table H.1. Percentage Difference in CMS Unweighted Mean Pre-Service and Immediate Post-Service Work Estimates and RAND Estimates, by Procedure Category (Chapter 5)

Procedure Category	CMS Work Estimate Mean RVUs	Typical RAND Predicted Work % Difference	All POS RAND Predicted Work % Difference
Total	1.48	–0.63	–0.60
CMS intra-service time categories			
0 to 30 minutes	0.58	5.50	4.96
31 to 70 minutes	1.15	0.86	0.45
71 to 120 minutes	1.79	–3.32	–3.35
Over 120 minutes	2.68	–1.12	–0.50
Global period			
0	0.97	–1.70	–1.91
10	0.78	–1.43	–1.56
90	1.77	–0.28	–0.20
Not applicable	0.62	–18.46	–17.37
Typical place of service			
ED	0.61	3.20	3.95
Inpatient	2.05	–0.70	–0.83
Office	0.63	2.20	2.45
Ambulatory facility (outpatient or ASC)	1.30	–1.18	–0.97
Risk category			
Office-based	0.54	–1.25	–1.32
ASC	1.25	–0.03	–0.07
Hospital outpatient	1.93	0.35	0.25
Inpatient only	2.59	–1.53	–1.37
Body system grouping			
Nervous system	1.98	–3.73	–3.70
Endocrine system	2.37	0.48	0.44
Eye	0.94	0.47	0.42
Ear	0.95	3.73	4.18
Nose, mouth, and pharynx	1.06	–0.27	–0.54
Respiratory system	1.62	0.48	0.49

Procedure Category	CMS Work Estimate Mean RVUs	Typical RAND Predicted Work % Difference	All POS RAND Predicted Work % Difference
Cardiovascular system	2.00	–0.47	–0.49
Hemic and lymphatic system	1.74	3.31	3.51
Digestive system	1.64	0.20	0.34
Urinary system	1.39	1.95	1.91
Male genital organs	1.27	2.44	2.66
Female genital organs	1.64	–0.71	–0.70
Musculoskeletal system	1.42	–0.86	–0.88
Integumentary system	1.04	–2.21	–2.07
Miscellaneous services	0.72	–5.49	–4.83
Number of annual Medicare procedures			
Less than 1,000	1.68	–0.95	–0.79
1,000 to 9,999	1.38	–0.80	–0.96
10,000 to 99,999	1.06	0.61	0.50
100,000 or more	0.66	11.48	10.66

Table H.2. Percentage Difference Between CMS Unweighted Mean Post-Operative Work Estimates and RAND Estimates, by Procedure Category (Chapter 6)

Procedure Category	CMS Work Estimate Mean RVUs	Typical Predicted Work % Difference	All POS Predicted Work % Difference
Total	4.23	–9.8	–5.5
CMS intra-service time categories			
0 to 30 minutes	0.47	–3.2	–3.6
31 to 70 minutes	2.13	–6.8	–7.0
71 to 120 minutes	5.39	–12.1	–6.6
Over 120 minutes	11.05	–10.0	–4.2
Global period			
0	0.03	–74.8	–100.0
10	0.93	–3.2	–6.2
90	6.17	–9.8	–5.3
Not applicable	0.00	N/A	N/A
Typical place of service			
ED	1.15	–21.4	–11.1
Inpatient	7.58	–6.9	–4.1
Office	0.88	–7.4	–5.9
Ambulatory facility (outpatient or ASC)	2.42	–18.8	–9.6
Risk category			
Office-based	0.59	–4.5	–3.8
ASC	2.57	–12.2	–9.4
Hospital outpatient	4.41	–15.7	–7.6
Inpatient only	11.11	–7.8	–2.7
Body system grouping			
Nervous system	5.53	–14.8	–3.7
Endocrine system	6.13	–23.1	–6.6
Eye	2.92	–6.4	–4.1
Ear	2.20	–16.5	–11.3
Nose, mouth, and pharynx	2.36	–14.5	–11.6
Respiratory system	5.30	–8.1	–2.9
Cardiovascular system	5.82	–3.0	–1.7
Hemic and lymphatic system	4.00	–6.5	–5.9
Digestive system	5.46	–3.6	–3.5

Procedure Category	CMS Work Estimate Mean RVUs	Typical Predicted Work % Difference	All POS Predicted Work % Difference
Urinary system	4.14	–19.0	–9.3
Male genital organs	2.96	–17.8	–4.1
Female genital organs	4.03	–20.2	–9.5
Musculoskeletal system	4.11	–11.6	–7.5
Integumentary system	2.32	–12.3	–9.8
Miscellaneous services	1.64	–1.7	–1.6
Number of annual Medicare procedures			
Less than 1,000	5.55	–10.6	–5.6
1,000 to 9,999	3.37	–8.1	–5.5
10,000 to 99,999	1.70	–7.2	–4.4
100,000 or more	0.59	–4.8	3.8

Table H.3. Percentage Difference Between CMS Unweighted Mean IWPUT Estimates and RAND Model Estimates (Chapter 7), by Procedure Category

Procedure Category	CMS Estimate Mean RVUs	Model 1a % Difference	Model 1b % Difference	Model 1c % Difference	Model 2 % Difference
Total	0.06	37.14	20.18	3.22	32.52
CMS intra-service time categories					
0 to 30 minutes	0.05	28.41	15.77	3.14	23.81
31 to 70 minutes	0.05	44.67	24.41	4.16	40.40
71 to 120 minutes	0.06	41.41	22.51	3.61	36.20
Over 120 minutes	0.07	25.00	13.06	1.12	20.32
Global period					
0	0.07	23.49	11.86	0.23	22.15
10	0.04	29.96	18.40	6.85	27.44
90	0.06	43.69	23.83	3.98	37.54
Not applicable	0.07	9.01	2.52	–3.97	6.10
Typical place of service					
ED	0.05	55.14	35.21	15.28	44.51
Inpatient	0.06	27.99	16.44	4.88	27.08
Office	0.05	41.58	23.89	6.21	34.63
Ambulatory facility (outpatient or ASC)	0.06	43.94	22.04	0.13	36.69
Risk category					
Office-based	0.05	25.12	13.70	2.28	21.08
ASC	0.05	44.65	23.56	2.47	39.90
Hospital outpatient	0.07	28.82	15.50	2.17	26.01
Inpatient only	0.06	27.02	16.42	5.82	21.91
Body system grouping					
Nervous system	0.06	32.15	17.22	2.28	23.70
Endocrine system	0.06	51.11	27.67	4.23	34.51
Eye	0.08	41.98	20.67	–0.65	37.33
Ear	0.07	25.50	13.16	0.82	22.72
Nose, mouth, and pharynx	0.04	39.44	21.80	4.16	31.45
Respiratory system	0.05	27.70	19.66	11.62	20.66
Cardiovascular system	0.08	18.44	11.10	3.75	18.10
Hemic and lymphatic system	0.06	21.52	12.56	3.61	22.93
Digestive system	0.06	38.47	21.48	4.48	38.89

Procedure Category	CMS Estimate Mean RVUs	Model 1a % Difference	Model 1b % Difference	Model 1c % Difference	Model 2 % Difference
Urinary system	0.07	34.32	18.16	2.00	26.04
Male genital organs	0.06	32.74	16.71	0.68	24.79
Female genital organs	0.06	32.25	17.16	2.06	24.21
Musculoskeletal system	0.05	46.80	25.29	3.77	41.44
Integumentary system	0.05	58.12	30.79	3.47	49.61
Miscellaneous services	0.05	30.93	14.58	–1.78	28.93
Number of annual Medicare procedures					
Less than 1,000	0.05	43.24	25.33	7.42	37.87
1,000 to 9,999	0.06	36.80	19.90	3.00	32.35
10,000 to 99,999	0.06	25.54	10.47	–4.61	22.51
100,000 or more	0.08	1.92	–9.96	–21.85	–0.36

Table H.4. Percentage Difference Between CMS Intra-Service Work Unweighted Mean Estimate and RAND Model Estimates, by Procedure Category (Chapter 7)

Procedure Category	CMS Estimate Mean RVUs	Model 1a % Difference	Model 1b % Difference	Model 1c % Difference	Model 2 % Difference
Total	4.84	9.83	–1.66	–13.15	6.17
CMS intra-service time categories					
0 to 30 minutes	0.84	29.24	17.22	5.19	25.89
31 to 70 minutes	2.52	18.68	2.97	–12.73	15.60
71 to 120 minutes	5.61	12.21	–1.55	–15.31	8.24
Over 120 minutes	12.85	3.49	–4.91	–13.30	–0.28
Global period					
0	2.53	2.77	–5.28	–13.34	2.02
10	1.52	5.13	–3.50	–12.13	5.18
90	6.12	11.38	–0.86	–13.09	7.07
Not applicable	5.12	–20.92	–19.50	–18.08	–16.44
Typical place of service					
ED	1.28	39.13	20.71	2.29	32.00
Inpatient	7.54	11.17	2.16	–6.85	7.03
Office	1.37	10.50	–4.95	–20.39	9.85
Ambulatory facility (outpatient or ASC)	3.75	6.54	–9.21	–24.97	3.45
Risk category					
Office-based	1.31	4.01	–5.00	–14.02	3.34
ASC	3.32	10.67	–4.90	–20.46	8.55
Hospital outpatient	7.09	14.34	2.84	–8.67	7.80
Inpatient only	10.78	8.85	0.71	–7.44	4.08
Body system grouping					
Nervous system	5.86	19.24	5.04	–9.16	8.47
Endocrine system	8.42	17.70	–0.51	–18.73	4.52
Eye	3.78	6.90	–10.21	–27.32	5.13
Ear	6.30	9.87	–0.09	–10.06	10.07
Nose, mouth, and pharynx	2.70	15.00	0.11	–14.78	11.47
Respiratory system	3.71	29.96	22.77	15.59	20.02
Cardiovascular system	9.53	0.69	–3.25	–7.20	–0.19
Hemic and lymphatic system	5.93	1.90	–4.65	–11.19	1.11
Digestive system	5.45	2.70	–8.08	–18.86	3.18
Urinary system	5.36	18.05	5.57	–6.92	10.49
Male genital organs	4.35	7.04	–6.75	–20.54	–0.83
Female genital organs	5.41	15.25	1.26	–12.73	6.98

Procedure Category	CMS Estimate Mean RVUs	Model 1a % Difference	Model 1b % Difference	Model 1c % Difference	Model 2 % Difference
Musculoskeletal system	3.62	14.06	–0.82	–15.70	10.10
Integumentary system	2.87	15.45	–3.95	–23.35	12.29
Miscellaneous services	2.38	–4.45	–11.46	–18.46	–1.06
Number of annual Medicare procedures					
Less than 1,000	5.53	15.24	3.44	–8.35	11.03
1,000 to 9,999	4.47	4.27	–6.82	–17.91	1.11
10,000 to 99,999	3.48	–4.79	–15.77	–26.75	–6.71
100,000 or more	1.91	–10.14	–20.29	–30.44	–12.16

Table H.5. Percentage Difference Between CMS Unweighted Mean Total Work RVUs and RAND Model Estimates, by Procedure Category (Chapter 8)

Procedure Category	CMS Estimate Mean Total Work RVUs	Model 1a % Difference	Model 1b % Difference	Model 1c % Difference	Model 2 % Difference	Model 3 % Difference
Total	10.47	0.71	–3.81	–8.33	0.61	–0.50
CMS intra-service time categories						
0 to 30 minutes	1.85	14.59	8.98	3.37	12.51	10.78
31 to 70 minutes	5.80	6.24	–0.48	–7.21	4.39	5.76
71 to 120 minutes	12.80	0.32	–5.14	–10.61	0.18	–0.48
Over 120 minutes	26.58	–2.72	–5.53	–8.35	–1.81	–4.36
Global period						
0	3.46	2.28	–5.43	–13.15	0.49	–0.76
10	3.19	1.37	–2.47	–6.31	0.02	0.73
90	14.07	0.71	–3.63	–7.97	0.76	–0.40
Not applicable	5.73	–20.61	–19.39	–18.17	–16.50	–19.61
Typical place of service						
ED	3.04	8.16	0.99	–6.18	10.13	5.69
Inpatient	17.16	2.16	–0.55	–3.27	1.29	–0.12
Office	2.80	3.71	–3.80	–11.31	3.67	2.78
Ambulatory facility (outpatient or ASC)	7.47	–3.26	–11.40	–19.53	–1.66	–2.04
Risk category						
Office-based	2.37	0.74	–4.07	–8.88	0.40	–1.68
ASC	7.13	0.69	–6.29	–13.28	0.43	–0.17
Hospital outpatient	13.43	3.64	–2.28	–8.20	2.57	2.66
Inpatient only	24.48	0.35	–1.95	–4.25	0.51	–1.11
Body system grouping						
Nervous system	13.31	1.86	–2.07	–6.00	1.56	0.73
Endocrine system	16.92	0.24	–7.10	–14.44	–0.12	0.97
Eye	7.57	0.97	–7.47	–15.92	0.94	–0.11
Ear	9.04	2.81	–3.43	–9.67	4.15	1.91
Nose, mouth, and pharynx	6.12	1.17	–4.92	–11.01	0.71	1.04
Respiratory system	10.63	7.53	5.03	2.53	6.23	–1.14
Cardiovascular system	17.36	–0.31	–2.56	–4.81	–0.50	–1.07
Hemic and lymphatic system	11.67	–0.64	–2.84	–5.03	–0.67	0.23
Digestive system	12.53	–0.04	–4.35	–8.65	0.05	–0.34

Procedure Category	CMS Estimate Mean Total Work RVUs	Model 1a % Difference	Model 1b % Difference	Model 1c % Difference	Model 2 % Difference	Model 3 % Difference
Urinary system	10.89	1.98	–2.63	–7.23	2.02	0.53
Male genital organs	8.57	–1.83	–8.06	–14.30	–1.45	–0.91
Female genital organs	10.93	0.23	–4.62	–9.46	0.05	0.06
Musculoskeletal system	9.14	0.37	–4.59	–9.54	0.41	–1.02
Integumentary system	5.99	1.88	–6.53	–14.95	1.06	1.04
Miscellaneous services	4.60	–4.81	–7.74	–10.68	–2.47	–13.07
2012 Medicare volume						
1st quartile	12.75	1.95	–2.17	–6.28	2.23	0.41
2nd quartile	9.18	–0.78	–5.68	–10.58	–1.48	–0.99
3rd quartile	6.07	–3.68	–9.99	–16.30	–4.59	–5.84
4th quartile	2.99	–1.52	–8.03	–14.54	–2.95	–5.44