Executive Summary

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

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

Published by the RAND Corporation, Santa Monica, Calif.

Since 1992, the Centers for Medicare & Medicaid Services (CMS) has used a system known as the resource-based relative value scale (RBRVS) to pay physicians and nonphysician practitioners for their professional and technical services. Each service is represented by a code using the Current Procedural Terminology (CPT) system.[1] The value of the service is measured in relative value units (RVUs). The total work RVUs estimated for a longer, high-risk, complicated procedure will be higher than those estimated for a shorter, low-risk, simpler procedure. To determine the fee schedule payment, RVUs are multiplied by a dollar conversion factor. Under the RBRVS, payment for a specific service is broken into three elements: physician work, practice expense, and malpractice expense.[2] The focus of this project is on physician work, which reflects the physician's effort in providing a service.[3]

Driven by concerns with the current process for valuing services under the RBRVS, the Affordable Care Act required that CMS establish a process to validate the physician work values assigned to services. CMS asked RAND to develop a model for valuing services that uses data from existing databases that are independent of the current valuation process. In response, RAND investigated the feasibility of developing such a model and the methodological issues and limitations involved in doing so. This RAND report is a freestanding executive summary of the RAND project reported more fully in a longer and more detailed report entitled *Development of a Model for the Validation of Work Relative Value Units for the Medicare Physician Fee Schedule*, by Barbara Wynn, Lane Burgette, Andrew Mulcahy, Edward Okeke, Ian Brantley, Neema Iyer, Teague Ruder, and Ateev Mehrotra (http://www.rand.org/t/RR662).

Concerns with the Current Process

There have been concerns raised about the current process used by CMS to value physician work. If a service is overvalued, an incentive may be created to provide unnecessary services; if a service is undervalued, the service may be hard to obtain, and there could be potential access problems. When determining the work value for a given service, CMS considers

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

[2] For most physician services, Medicare pays 80 percent of the fee schedule amount, and the beneficiary is responsible for the remaining 20 percent. The fee schedule amount is geographically adjusted. Other adjustments are also applied for practitioner type (for example, decreased payment for nurse practitioners) and type of service (for example, primary care furnished in health shortage areas).

[3] 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.

recommendations from a committee of the American Medical Association/Specialty Society Relative Value Scale Update Committee (RUC). The RUC is composed of representatives of physician specialty societies. The Medicare Payment Advisory Commission (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 (MedPAC, 2011).

The RUC's process to determine how much work is involved for a given service relies on physician surveys conducted by the specialty societies. The surveys ask the physicians to subjectively estimate the total work relative to another procedure and the time required to perform the procedure. Related concerns are the potentially small number of respondents to the surveys, low response rates, and the unclear generalizability of the responses.[4] These concerns have the potential to introduce biases in the survey results. For example, with respect to intra-service time (that is, the time required to perform a given service), the times estimated in the physician surveys consistently exceed the times seen in external data sources (McCall, Cromwell, and Braun, 2006; Rich, 2007; Cromwell et al., 2010). These concerns have prompted calls for a more objective assessment of work that uses external data sources instead of physician surveys.

Overview of RBRVS

Under the RBRVS, each service's work is measured in RVUs. RVUs are converted into a dollar payment amount using a conversion factor. For example, if a procedure has a value of ten physician work RVUs and the conversion factor for 2014 is $35.822, the payment for physician work for that procedure is $358.22.

Under the RBRVS, physicians receive a single payment for their effort in providing a procedure based on total work RVUs.[5] Surgical procedures generally have a 0-, 10-, or 90-day global period. The total work RVU payment for the global period means that there is no separate payment for follow-up post-operative visits during that global period.[6]

Total physician work can be broken down into four components: (1) pre-service work (for example, positioning prior to surgery), (2) intra-service work (the performance of the procedure or "skin-to-skin time"), (3) immediate post-service work (for example, management of a patient in the post-operative recovery room), and (4) post-operative E&M visits (only applicable for

[4] As described in the full report, the RUC has taken steps to address some of these concerns.

[5] In this report we focus on physician work. Total payment also includes practice expense and professional liability insurance.

[6] A 90-day global period also includes the day before the procedure.

procedures paid on a global period).[7] One can calculate total work by summing each of the four components together, which has been termed the *building block method* (BBM).

Total RVUs = pre-service + intra-service + immediate post-service + post-operative E&M visits

Each of these subcomponents can be further broken down into time multiplied by intensity (also called *work per unit time*). Intensity reflects the physician's cognitive effort and judgment, technical skill and physical effort, and stress due to potential patient risk. For example, intra-service work can be divided into intra-service time and intra-service work per unit time (IWPUT).

Intra-service work RVUs = intra-service time × intra-service work per unit time (IWPUT)

An example might help make this breakdown more concrete. In Table 1, we show the valuation of CPT 33510 (coronary artery bypass graft, one vein graft). The procedure is valued at 35.0 total work RVUs. We can break down the 35.0 RVUs[8] assigned to this service into pre-service work (1.8 RVUs), intra-service work (14.9), immediate post-service work (0.9), and post-operative E&M visit work (17.3).[9] Intra-service work can be broken down into time (154 minutes) multiplied by intensity (0.097). We return to this example later in this report.

Table 1. Illustration of Valuation of a Surgical Procedure, CPT 33510 (Coronary Artery Bypass Graft, One Vein Graft)

Work Component	Work (RVUs)	Minutes	Intensity
Pre-service (evaluation, positioning, scrub)	1.8	95	0.019
Intra-service	14.9	154	0.097
Immediate post-service	0.9	40	0.022
Post-operative E&M visits	17.3	429	0.040
Total	35.0	718	0.049

[7] Post-operative E&M visits are only included for procedures that are paid on a global period in which a single payment covers the work of practitioners for a 10- or 90-day period around the performance of the procedure. For 0-day global periods, visits performed on the day of the procedure (typically preoperative) are included in the payment. Visits that are not included in the global payment are separately payable. In its final rule for the 2015 physician fee schedule, CMS indicated that it would begin transitioning the 10-day and 90-day global periods to 0-day global periods in 2017. As noted below, the RAND model could inform the valuation of surgical procedures during this transition.

[8] Components do not total 35.0 because of rounding.

[9] We use a "reverse" BBM to break down the total work RVUs. We can estimate the RVUs for each component other than intra-service work by multiplying CMS time estimates by standard intensity factors. We subtract the RVUs for these components from the total work RVUs to derive an estimate for intra-service work. Intra-service intensity can be calculated by dividing intra-service work by intra-service time.

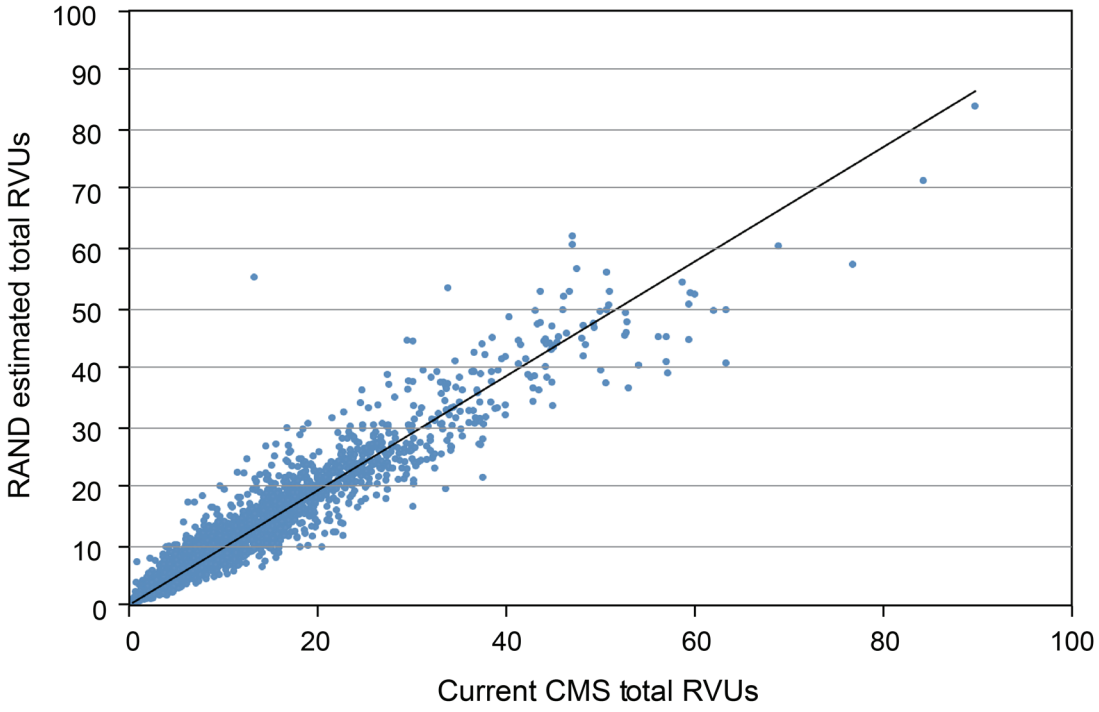
Project Objectives and Overview of Model and Results

RAND's objective was to develop models to generate independent estimates of total work and work subcomponents: intra-service time, intra-service intensity, pre- and immediate post-service work, and post-operative E&M visits. For this project, we focused on 3,179 surgical procedures, including (1) surgery where there is an incision of some kind; (2) other types of invasive procedures, such as colonoscopy, that do not require an incision; and (3) selected medical procedures, such as interventional cardiology, that often are performed in an operating room setting. We excluded surgical procedures rarely performed in the Medicare population or almost always performed in office settings. We did not focus on other parts of the RBRVS system, such as E&M visits or the interpretation of radiology or pathology specimens. Our procedure selection was based on feasibility. The time required to perform the procedure, intra-service time, is a critical piece of information in developing our models. There were no publicly available datasets available that provide necessary time data for nonsurgical services or office-based procedures.

As described in more depth below, using external data sources, RAND determined characteristics of each surgical procedure, such as intra-service time, years of training among physicians who perform the procedure, and the mortality risk after the procedure. Based on these characteristics, RAND's model estimates for each procedure the number of RVUs for total work and the four components of work.

On average, RAND's model results are highly correlated with current CMS estimates. This is illustrated in Figure 1, which compares the current total work RVUs with RAND estimated total work RVUs. There are procedures, however, where there is a large difference between current CMS values and the RAND model results. These procedures might be further investigated to determine if they are misvalued currently.

Figure 1. Comparison of Current Total Work RVUs with RAND Estimated Total Work RVUs

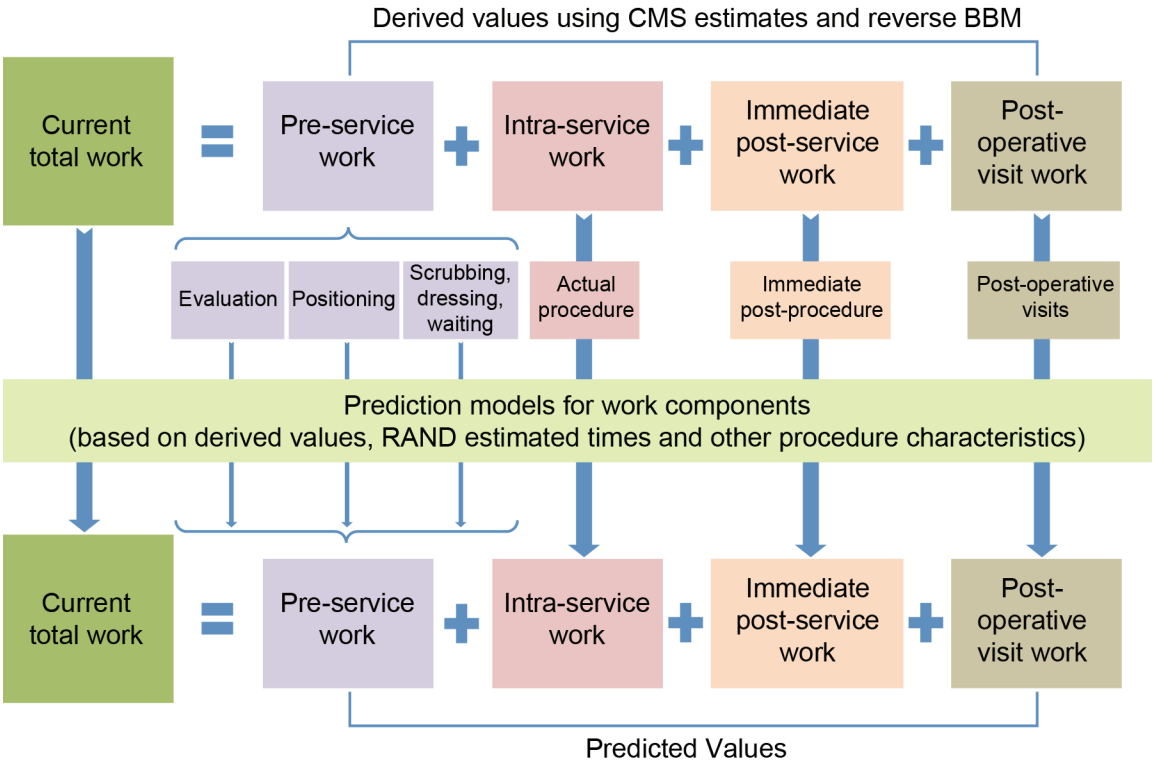


NOTE: Results are from RAND Model 1a.

Developing a model to predict work for a surgical procedure is complicated, and we are cognizant that there is no single optimal approach. Therefore, RAND researchers generated different models (Models 1a, 1b, 1c, 2, and 3) that reflect key methodological decisions and trade-offs. These methodological decisions are detailed below.

All of the models except for Model 3 are estimated using the BBM. In the BBM approach, there are separate models for each component, and these are added together to estimate total work (Figure 2). Under Model 3, we use a single prediction model to estimate total work directly.

Figure 2. Overview of Building Block Method Approach and Single Prediction Model Approach



An illustration of the model results might help to make this more concrete. CPT 33510 (coronary artery bypass with single vein graft) has a 90-day global period. Some of the characteristics we capture are intra-service time (based on external databases, 162 minutes), the rate for complications in subsequent 30 days (36 percent), malpractice risk (based on the premiums for the specialties performing the service), and the least-resource-intensive setting in which Medicare covers the service (inpatient). (The full set of characteristics is discussed below.) We use the values for these and the other characteristics in a prediction model to estimate the total work and work components (Table 2). For this particular procedure, the RAND model estimate and the current CMS estimates are very similar. However, for many procedures there is a notable difference. For example, for CPT 45380 (colonoscopy with biopsy), the RAND model estimate is 26 percent lower than the current CMS estimate for total work RVUs (Table 2).

Table 2. Illustration of Model Results for Two Procedures

CPT 33510: Coronary Artery Bypass Graft (CABG) with Single Vein Graft

Work Component	Current RVUs	RAND Model RVUs
Total	34.98	34.17
Pre-service and immediate post-service	2.74	2.58
Post-operative E&M visits	17.30	17.23
Intra-service	14.94	14.36
Intra-service time	154 min.	162 min.

CPT 45380: Colonoscopy with Biopsy

Work Component	Current RVUs	RAND Model RVUs
Total	4.43	3.26
Pre-service and immediate post-service	1.50	0.94
Post-operative E&M visits	0	0
Intra-service	2.93	2.33
Intra-service time	52 min.	17 min.

NOTE: Results are from RAND Model 1a.

Key Methodological Steps

Characteristics of Procedures

We measure 16 characteristics of the procedures, including intra-service times and measures for procedure complexity and risk and patient complexity that we use in the models (Table 3).

Table 3. Summary of Procedure Characteristics Used in Prediction Models

Characteristics	Brief Explanation
Intra-service time	How long it takes to perform the service, “skin-to-skin” time. Key variable in models
Code grouping	Codes grouped by a combination of clinical characteristics and the amount of work required (for example, laser eye procedures)
Body system	Codes grouped by body system (for example, ear procedures)
Global period	Most procedures have a 0-, 10-, or 90-day global period in which all post-operative care is included in single payment
Risk level	Categorical variable based on the least-resource-intensive setting in which Medicare covers the procedure
Laparoscopic or thoracic procedure	Included because these may have unique aspects that impact pre-service and immediate post-service work
Comorbidities	Captured by the count of comorbidities
Length of stay	Median length of stay in a hospital setting among those who receive the service
Intensive care unit (ICU) days	Median length of stay in an ICU among those who receive the service
Age	Average age of Medicare beneficiary receiving the service
Gender	Proportion of Medicare beneficiaries who receive the service who are female
Major complications	Proportion of Medicare beneficiaries who receive the service who have a major complication in the subsequent 30 days
Mortality rate	Proportion of Medicare beneficiaries who receive the service who die in the subsequent 30 days
Malpractice risk	Calculated for each specialty and individualized for a given service based on mix of specialties that bill for that service

Characteristics	Brief Explanation
Years of training	Calculated for each specialty and individualized for a given service based on mix of specialties that bill for that service
Urgency of decisionmaking	As a marker for urgency, what fraction of the procedures performed for Medicare beneficiaries occur in the emergency department or on the first day of a hospitalization and admitted via the emergency department

NOTE: Not all characteristics were included in each prediction model. For example, malpractice risk was only included in models for intra-service work and IWPUT.

These characteristics were chosen based on two considerations. Based on prior research or theory, we believe that they should be related to physician work. For example, the initial developers of the RBRVS conceptualized intensity as (1) technical skill and physical effort, (2) mental effort and clinical judgment, and (3) psychological stress and risk. We used years of training as a proxy for technical skill. We used urgency of medical decisionmaking and malpractice risk as proxies for psychological stress and risk. The other key consideration was the ability to measure the characteristic in the datasets available to us.

Except for intra-service time, the characteristics of these procedures are measured using data from the Medicare population. The focus on just the Medicare population might be a limitation for Medicaid and other non-Medicare payers that also use the RBRVS to determine payments.

Determining RAND Intra-Service Time

One key procedure characteristic in our models is intra-service time ("skin-to-skin" surgical time). Its importance is illustrated by an extremely high correlation between intra-service time and total work.[10] There is no single data source that has intra-service time for all surgical procedures. We therefore use two different publicly available data sources to estimate surgical times. We focus on readily available data sources because they allow transparency and would be available to CMS for potential future updating of the model. We use Medicare anesthesia claims to estimate times for surgical procedures that require anesthesia. Because many surgical procedures typically do not require anesthesia, we also use data from the New York Statewide Planning and Research Cooperative System (SPARCS). It collects patient-level detail on the operating room time for every ambulatory surgery performed in a hospital outpatient department or ambulatory surgery center in New York State. If two or more surgical procedures are performed at the same time, it is unclear how long it took to perform each procedure. Therefore, in both databases we only use time data on single procedures performed in an operating session.

[10] The correlation coefficient between total work and intra-service time in CMS time file for the core procedures is 0.91.

As detailed in the full report, the time estimates for individual surgical procedures from Medicare anesthesia and SPARCS were consistent.

Medicare anesthesia claims are billed in fractions of 15-minute increments. We link how much time was billed for anesthesia with the claim for the relevant surgical procedure. This allows us to estimate the median anesthesia time for a given surgical procedure. Billed "anesthesia time" captures a different time component than "operating room time" available in the SPARCS data. Both times are different than the key variable of interest: intra-service or "skin-to-skin" time. To address these differences, we use a formula to transform anesthesia time and operating room time into surgical time. Our transformation is built on prior research by Silber et al. (2007, 2011) that studied the feasibility of using Medicare anesthesia data to estimate anesthesia and surgical times that were manually abstracted from patients' charts. They found that average anesthesia claim times were highly predictive of surgical times. We expand this transformation to operating room time and a broader set of surgical procedures.

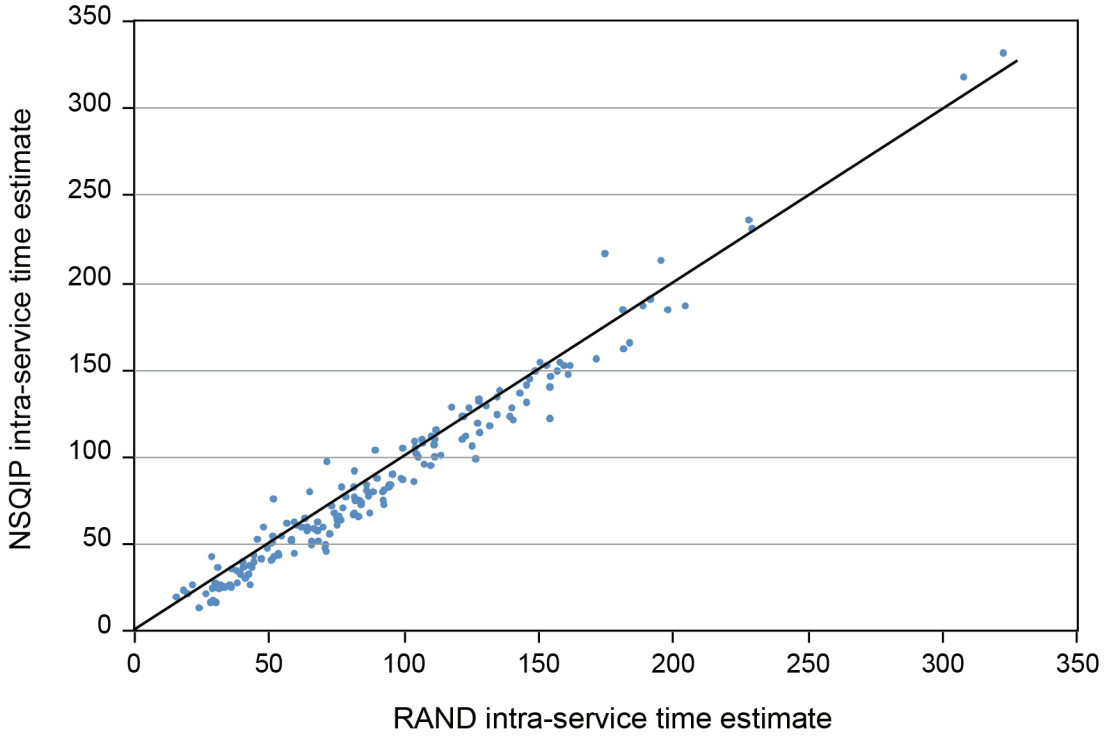
Another key issue is that some of the procedures in the independent databases we use (Medicare and SPARCS) were performed infrequently, and therefore we cannot precisely estimate how long a given procedure takes. There can be a small number of procedures because the procedure is rarely performed in the Medicare population, the procedure is almost always done in conjunction with other surgical procedures (and we only include single procedures), or the procedure is infrequently performed in a non-office setting. One mechanism to address this issue of varying amounts of available data is to use a type of statistical method called 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. We use the existing CMS estimates (mostly derived from the RUC's specialty society physician surveys) as this prior estimate.[11] Hence, if there are only a few observations for a procedure in the external databases, our estimates of their times will be close to the CMS estimates. For procedures that have a large number of observations in our database, however, the prior information (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 described in more depth below and consistent with prior work, the RAND estimates of intra-service time for 83 percent of surgical procedures are shorter than current CMS estimates. For example, for ambulatory procedures done with anesthesia, the RAND time is 10 percent shorter on average. To validate the RAND estimates of intra-service time, we compared the time estimates to the intra-service time using several well-established data sources that are not

[11] As detailed in the full report, we apply an adjustment factor to the CMS estimate before using it as a prior estimate. This adjustment reduces to the CMS time estimate by a multiplicative constant. We apply this adjustment factor to address the issue that CMS time estimates tend to be longer than observed surgical times. If we did not apply this adjustment factor, the time estimates for procedures with few available time observations might be too long.

publicly available or comprehensive. The first is the American College of Surgeons National Surgical Quality Improvement Program (NSQIP) data. In this database, individual hospitals across the country collect a large set of data on specific surgical procedures, including intra-service time. We find that the mean RAND intra-service time is highly correlated with the intra-service time that comes from the NSQIP data (Figure 3). Comparisons with our other databases found similar results.

Figure 3. Comparison of RAND and NSQIP Intra-Service Time Estimates[12]



Building Prediction Models

Using regression analysis to create a prediction model ideally would be built on a "gold standard" set of values for total work or the individual components (pre-service, intra-service, post-service, or post-operative E&M visits). We use the term "gold standard" to refer to a set of values for which there is agreement that these values are accurate. One challenge that we face is the lack of a gold standard for both total work and the individual work components. Therefore, we rely on current CMS estimates in building our prediction models. This means that, to the extent that there are systematic biases in the CMS estimates, we build those biases into our predicted values, and our models do not address the underlying problem. Post-operative E&M

[12] This comparison was limited to procedures with sufficient sample size of 250 cases in both the NSQIP data and the databases RAND used to estimate time. NSQIP intra-service time is measured by the median. The RAND intra-service time is for typical POS.

visits are a good example of this issue. We can identify procedures where the current CMS estimates are inconsistent with the characteristics of the procedure. For example, we might find a procedure where the current CMS estimate is five visits, but based on the characteristics of the procedure, we estimate seven visits. However, as detailed above, there is concern that the number of post-operative E&M visits is systematically overestimated *across all codes*. Our RAND model for post-operative E&M visits would include this systematic bias in its estimates. While this limitation is important, the RAND model retains important advantages over the current valuation process. These advantages are detailed in the "Strengths and Weaknesses of RAND Model" section below.

Our starting point in deriving the initial values for the RAND model is total work RVUs in current CMS estimates. Arguably, the total work RVUs are the most accurate estimates in the current process because they receive more scrutiny than the individual components.[13] These total work RVUs are what we use to build our single prediction model. For the other models, we use the CMS time estimates and standard intensity values for pre-service work, immediate post-service work, and post-operative E&M visits to build the prediction models for these work components. There is no standard intensity estimate for intra-service work. We therefore use what is commonly referred to as a reverse BBM to derive the estimates for intra-service work. Under the reverse BBM, we start with the total work RVUs and subtract the RAND prediction model estimates for pre-service work, immediate post-service work, and post-operative E&M visits to generate an estimate of intra-service work.[14]

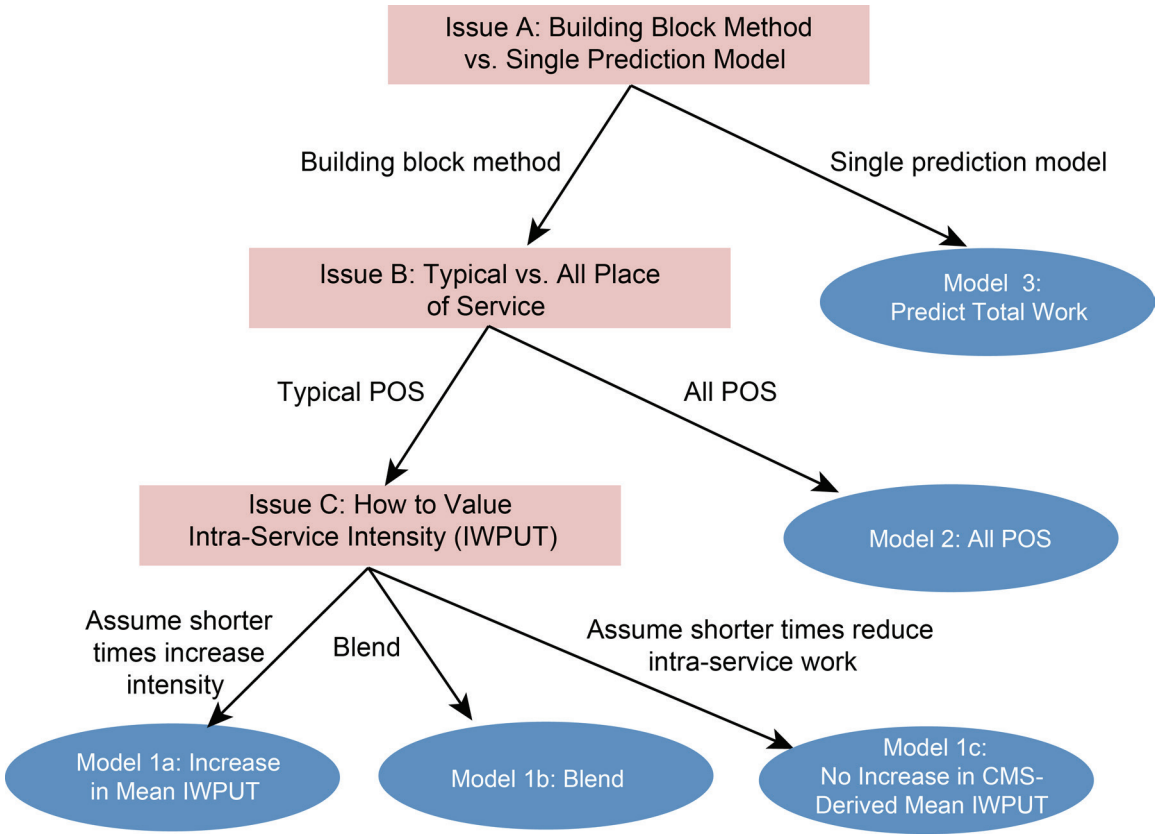
Differences Across RAND Models

As noted above, building a prediction model can be complicated, and there are many options we could pursue. Figure 4 summarizes the major issues that we considered in modeling total work and how we combine choices on these options into models. We have chosen these models to highlight the impact of different modeling choices. Alternative models could be created by combining the choices differently.

[13] 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.

[14] In the reverse building block approach, intra-service work = (total work) – (pre-service work) – (immediate post-service work) – (post-operative E&M visit work).

Figure 4. Overview of RAND Model



Issue A: Building Block Method Versus Direct Estimation of Total Work

Models 1and 2 use the BBM, while Model 3 estimates total work RVUs directly using a single prediction model. The direct estimation of total work is consistent with the RUC valuations, which are for total work RVUs only and not the individual work components.[15] Despite the simplicity of estimating total work directly, there are several potential issues with this approach. Possibly the most important is that Model 3 does not explain why there might be a difference in total work between the RAND model and CMS estimates. In contrast, the models that use the BBM approach provide insight on which individual components may be misvalued.

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

Most procedures are performed in more than one clinical setting or POS (for example, inpatient, outpatient hospital, emergency department, office). The times for the work components and the number of post-operative E&M visits vary across POS. The surveys used in the RUC process ask physicians to consider the "typical" POS when valuing a service. Having external data available to estimate surgical times and other characteristics begs the question of whether a procedure should be valued based on the typical POS (where it is performed most often) or whether the

[15] While the surveys collect information on the individual components that inform the valuation, the valuation is for relative total work.

values should reflect the full range of POS in which the procedure is furnished, which we term the "all POS" approach. Models 1 and 3 maintain the current RUC framework for valuing surgical procedures by using the "typical" POS. The disadvantage is that this framework does not reflect the reality that many procedures are done in multiple settings. Also, the distribution of sites of service might have changed a great deal since the most recent valuation by the RUC, as there has been a general shift in surgery to more outpatient procedures. The all POS approach used in Model 2 potentially provides a more accurate valuation for the procedures, as it reflects all sites where it is performed.

Issue C: How to Value Intensity

As discussed earlier, intra-service work can be derived as the product of intra-service time and intra-service intensity (IWPUT). The RAND time estimates are systematically shorter than the CMS time estimates. The core question underlying Issue C is the extent to which the differences between RAND and CMS times affect IWPUT and/or intra-service work. In other words, if it takes less time to perform a procedure, does the intensity of performing the procedure (IWPUT) increase? Or should the shorter time required to perform a procedure have no impact on IWPUT and therefore decrease intra-service work? One argument for maintaining the current average IWPUT value is that it reflects a core level of intensity for surgical procedures. It is notable that the IWPUT for surgical procedures has not changed over the last 20 years. In the original development of the RBRVS, the average IWPUT for surgical procedures was 0.057, and the average IWPUT derived from the current CMS estimates for surgical procedures is 0.056. It could even be argued that on average IWPUT should have decreased over the last two decades. Mortality after surgical procedures has progressively declined, and improved technology and efficiency gains suggest that many 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 the average IWPUT for surgical procedures. The first is that the current RUC process for valuation focuses on total work. The time estimates for the individual components do not receive as much scrutiny. Intra-service work and IWPUT are derived values, and inaccuracies in other work subcomponents (in particular, post-operative E&M visits) could make the intra-service work values suspect. It is possible that IWPUT is underestimated in the CMS estimates, particularly for procedures with 90-day global periods. Another argument for increasing IWPUT is that while technology and efficiency gains have reduced average surgical procedure times, the work per unit time may be correspondingly higher. For example, while a colonoscopy can be done more quickly, IWPUT during colonoscopy may be higher because physicians must maintain a higher level of mental focus and use higher levels of technical skill.

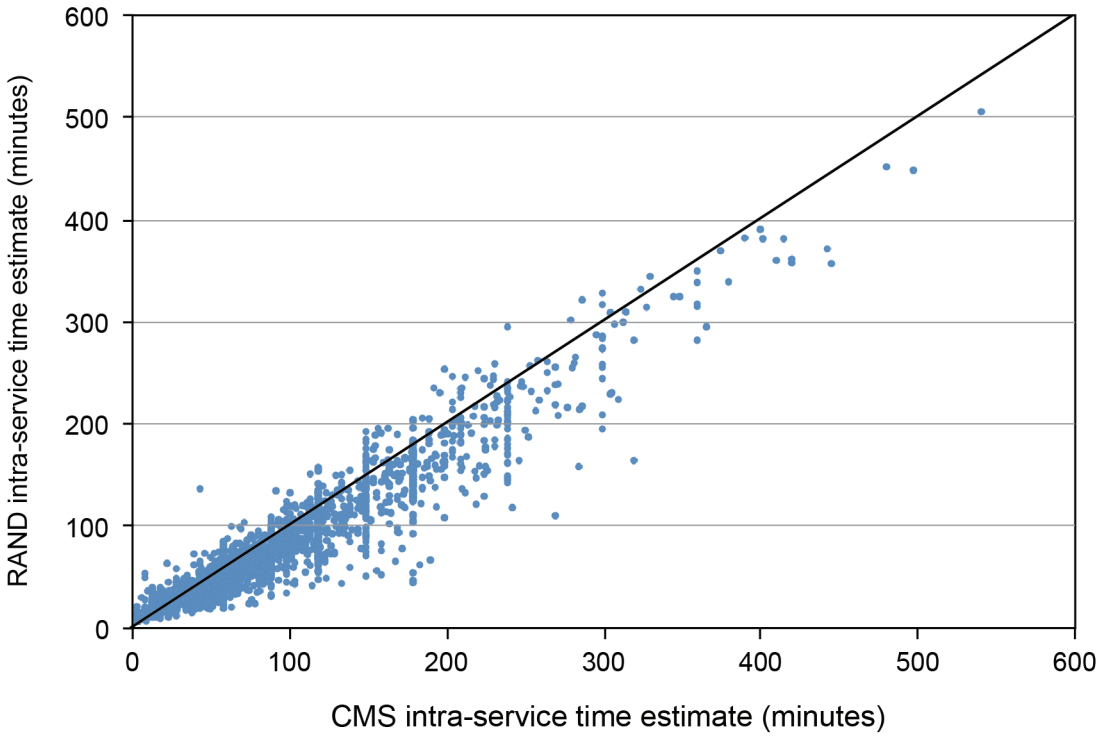
Issue C has critical implications for valuing intra-service work and therefore total work. However, the lack of a gold standard for intensity means that it is not possible to empirically assess which argument is correct. Instead, our Model 1 alternatives illustrate the range of

arguments and help inform the empirical impact of applying different arguments. Model 1a assumes that a shorter procedure time translates into higher intensity 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. 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.

Key Findings

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 (Figure 5). For 83 percent of the procedures, the RAND time is shorter than the CMS estimates. These results are consistent 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). This is a key finding because intra-service time is the key driver of total work.

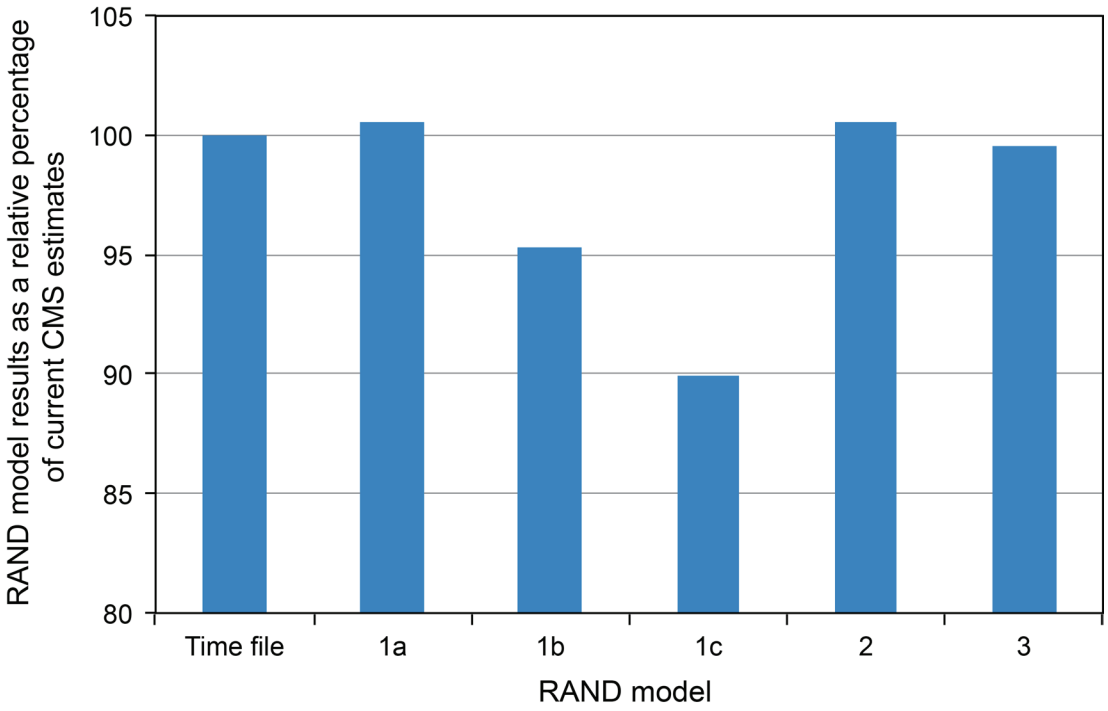
Figure 5. Comparison of RAND and CMS Intra-Service Time Estimates



NOTE: The RAND estimate is from all places of service. The RAND time estimate from typical places of service is very similar.

2. Average total work in RAND models is similar to CMS estimates, but there are important differences for some procedures. As demonstrated in Figure 6, the average total work in RAND Models 1a, 2, and 3 and CMS total work are nearly identical. Reflecting the reductions made in intra-service work for shorter intra-service times, the average RVUs for Models 1b and Model 1c are 4.8 percent and 10.0 percent lower, respectively, than the CMS average.

Figure 6. Average Total Work RVUs Predicted by Models Relative to CMS Values



NOTE: Results in Figure 6 are not weighted by volume of procedures.

While on average the valuations of surgical procedures are similar in Model 1a 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 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 4).

On average, across all surgical procedures, Issue A (BBM or single prediction model) does not make a difference (Model 1a versus Model 3). Nor does Issue B (typical POS or all POS) make a difference (Model 1a versus Model 2). In contrast, Issue C (how to value intensity) is important. Because Model 1c assumes that the average IWPUT has not changed, intra-service work is reduced substantially, and this drives the 10-percent reduction in total work. This is discussed in more depth below.

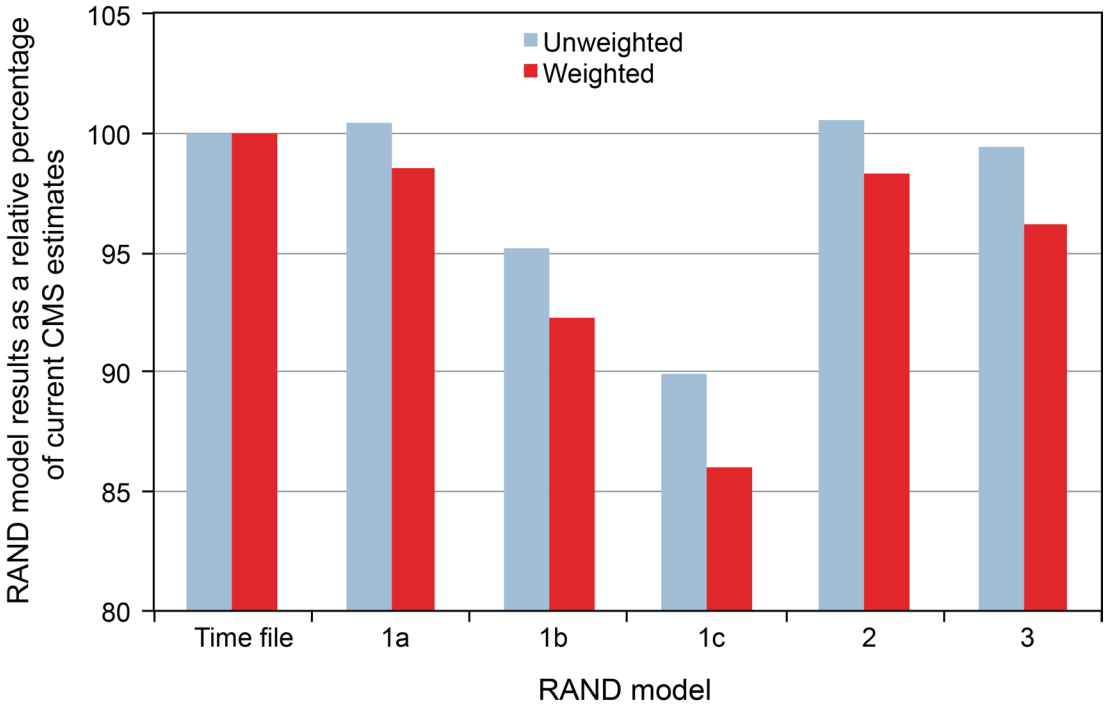
Table 4. Percentage Difference Between CMS Total Work RVUs and RAND Estimates, Unweighted by Procedure Volume, by Procedure Category

	CMS Estimate Mean Total RVUs	Model 1a % Difference	Model 1b % Difference	Model 1c % Difference	Model 2 % Difference	Model 3 % Difference
All procedures	10.54	0.5	–4.8	–10.0	0.6	–0.5
Intra-service time[16]						
0 to 30 minutes	1.89	13.9	8.5	3.2	12.1	10.3
31 to 70 minutes	5.80	5.8	–1.0	–7.85	4.3	5.4
71 to 120 minutes	12.80	–0.2	–6.2	–12.3	0.4	–0.7
Over 120 minutes	26.58	–2.6	–6.6	–10.7	–1.9	–4.0
Global period						
0 days	3.53	0.9	–4.9	–10.7	0.0	–1.7
10 days	3.22	1.15	–2.9	–7.0	0.3	0.5
90 days	14.07	0.6	–4.7	–10.1	0.7	–0.3
2012 Medicare volume						
<1000	12.76	1.9	–3.3	–8.4	2.3	0.5
1,000 to 9,999	9.22	–1.0	–6.4	–11.7	–1.6	–1.1
10,000 to 99,999	6.24	–4.5	–10.7	–16.8	–4.8	–6.3
100,000 or more	3.16	–4.6	–10.8	–16.9	–4.4	–7.6

[16] Procedures categorized by time in current CMS estimates.

3. The difference in total RVUs across RBRVS is greater than the average impact 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 more influence).[17] The difference between unweighted and weighted results is important because the weighted estimates capture what would be paid by Medicare. 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). 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.

Figure 7. Average Total Work RVUs Predicted by Models Relative to CMS Values, Unweighted and Weighted by Medicare Volume



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 inpatient E&M visits are included in the global period, but the procedure is typically performed outside the hospital. These corrections reduced the unweighted post-operative work RVUs by 10 percent.

[17] When we build our prediction models, each procedure code contributes equally in estimation.

5. The difference between the CMS estimates and the RAND estimates for IWPUT and intra-service work varies across the models. As noted above, the 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, lower intra-service work stays constant, intra-service time goes down, and therefore IWPUT increases (Table 5). Under Model 1c, IWPUT stays the same, intra-service time is lower, and therefore intra-service work is reduced.

One concern that has been noted with the current CMS estimates of IWPUT and intra-service work is that many values are nonsensical. For example, for some procedures, the derived estimates of intra-service work and IWPUT are negative. To help quantify the impact of the RAND model on nonsensical values, we created an IWPUT range. The bottom of the range was 0.0224, which is the IWPUT for the pre-service evaluation and positioning, and the top of the range is 0.11, which is the current IWPUT in the CMS estimates for a lung transplant with bypass (Table 5). Across the RAND models, there are many fewer codes with IWPUT values outside this range than in the current CMS estimates.

Table 5. Average IWPUT and Intra-Service Work in RAND Models Compared to CMS Estimates

	Mean IWPUT (RVUs/minute)	Percentage of IWPUT Values That Are Outside of Range[18]	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

Strengths and Weaknesses of RAND Model

Many of the key advantages of the RAND model address concerns with the current system for valuing physician services (Table 6). The RAND model's valuation of a given procedure is done in a transparent, consistent manner across all procedures using procedure time data from databases that are independent of the current valuation process. This helps address concerns that the current process is subject to bias and that revaluation of services focuses primarily on

[18] To address nonsensical IWPUT values, we created a floor equal to the standard IWPUT for pre-service evaluation and position, which is 0.0224. Using this floor in our prediction models contributes to higher IWPUT values and explains the increase in IWPUT values in Model 1c relative to the CMS time file.

undervalued services. Because the RAND models can be run in an automated manner, they can be run frequently, even yearly, and therefore incorporate efficiency gains and shifts in care. For example, if the time required for a procedure decreases or the location where the procedure performs shifts from an inpatient setting to an outpatient setting, then this would automatically be incorporated into the valuation of a procedure.

Table 6. How RAND Model Could Address Concerns with Current System

Concern with Current System as Described Above	How RAND 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 *all* surgical procedures on a frequent basis, even yearly. It can also 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 directly use data from physician surveys. (As noted in limitations, our models are built on CMS estimates and thus may incorporate their systematic biases.)
Derived intra-service intensity values are sometimes nonsensical	Intensity values that arise from RAND validation models result in no negative values or values that are so low that they lack face validity.
CMS may be overpaying for post-procedure care in the global period	RAND validation models only partially address this issue by making a correction when it appears there are too many inpatient E&M visits.
RUC process does not adequately address efficiency gains	Because RAND validation models can be run on a frequent basis for every procedure, efficiency gains can be incorporated regularly.

There are several key weaknesses of the RAND model. Possibly the most important is that, except for intra-service time, we lack a gold standard for total work or subcomponents of work. The RAND model therefore depends on current CMS estimates in building the prediction models. As we discuss above, this means that the valuations of procedures by most RAND models are, *on average,* similar to current estimates. Another way of framing this concern is that the RAND model simply shifts work from one procedure to another and help identify procedures in which the current CMS estimates are inconsistent with the characteristics of the procedure. The major issue is that if there are systematic biases in the CMS estimates, the RAND model builds those biases into our predicted values.

The lack of a gold standard is particularly important in relation to intensity. As described above, one key methodological issue is how shorter intra-service times affect IWPUT and/or intra-service work. The modeling decision on this issue could have significant impact on the valuation of surgical procedures and a 10-percent difference in total work across the models. As noted below in future steps, we believe that collecting independent intra-service work and intensity can help inform this decision.

To estimate how long a procedure takes—the intra-service time—RAND uses data from independent databases that capture the time on tens of thousands of procedures. We have validated the RAND time estimates and demonstrated that, *on average*, they are consistent with other well-established estimates of time. We also have demonstrated that, *on average*, the RAND time estimates are more accurate than the values in the current CMS valuation. However, for specific procedures, the RAND time estimates are possibly erroneous. For example, the RAND time estimates are based on the performance of a single procedure, but for some ophthalmology procedures the time estimates used in the valuation of a procedure assume that multiple procedures are done within a 90-day period.[19] There is also some confusion regarding whether the delivery of conscious sedation for a gastroenterological endoscopy procedure is included in the intra-service time. These examples emphasize the need for further refinement and validation of the RAND time estimates. The addition of time data in other databases would also help. As noted above, the intra-service time is by far the most critical variable in the valuation of a procedure.

The RAND model uses a common set of procedure characteristics across all surgical procedures to value work. Another limitation is that unique clinical issues specific to a procedure or specialty may not be captured in the characteristics we use. For example, emergent tracheostomy is a procedure only done in clinical situations in which seconds might make the difference between life and death. This characteristic of emergent tracheostomy is not fully captured in the RAND model. The RAND model captures the risk of a set of common major complications after a procedure, such as heart attack or blood clot. However, the model does not capture the risks of specific procedures, such as significant cosmetic defects or disabilities—for example, blindness. These risks might increase the intensity of a procedure, but the RAND model does not capture this risk. It is therefore likely that for a fraction of procedures where the valuation between the RAND estimates and current CMS estimates are significantly different, there is a clear clinical reason for why the RAND estimates are different. As detailed below, we believe a key future step is to obtain clinical input on potential factors to add to the RAND model.

[19] For example, the valuation of CPT 67228 (laser photocoagulation of diabetic retinopathy) assumed that more than one procedure was performed during the 90-day global period.

Applications of the RAND Model and Future Steps

We believe that CMS could use the RAND model estimates in two ways to validate the valuation of surgical procedures. First, the model estimates could be used to identify potentially misvalued codes. Comparing RAND estimates and current CMS estimates will identify services where the valuation is inconsistent with the characteristics of the service. Second, the RAND estimates might serve CMS as a useful counterpoint in assessing the RUC's valuation of a service. In some cases, a comparison of CMS/RUC estimates and RAND estimates will identify a clinical rationale for why a code is valued differently, and the CMS/RUC estimate may be more appropriate. In other cases, the RAND validation model results will highlight when a code is misvalued. 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 RAND models. 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 that have large discrepancies and therefore are potentially misvalued. The output for each work component from Model 1b could be used to identify which work component might be contributing to the potential differences in valuation.

In addition to providing an independent estimate for new, revised, or potentially misvalued codes, the model could be used to provide estimates for policy changes. For example, CMS recently announced its intention to phase out the 10-day and 90-day global periods for surgical procedures beginning in 2017. Because the RAND model provides independent estimates of post-operative visit work, the model could be adjusted to take this policy change into account and estimate RVUs for surgical procedures with no post-operative visit work included.

In their current form, we do not believe that the RAND model should replace the current valuation process. To refine the RAND model for surgical services, we recommend three steps. The first is to generate estimates for intra-service work for a small set of surgical services outside the RUC process. This would address a key limitation of the RAND model 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. These values for intra-service work would be used to calibrate the RAND prediction model. The RAND prediction model could then be used to estimate work RVUs for the full set of surgical procedures. The second step would be to improve the RAND time estimates by obtaining more data on the times for procedures, in particular those done in an office setting. Third, clinical input can be used to add or refine the procedure characteristics used in the RAND model.

The current RAND model focuses only on surgical procedures. Surgical procedures are only part of the RBRVS. E&M visits, interpretation of laboratory, pathology specimens, and radiology are all key parts of the RBRVS not included in the model. Significant effort will be necessary to

develop new models for the nonsurgical aspects of RBRVS. The key will be to identify external datasets with intra-service times for these services. Possibly the research being conducted by the Urban Institute on the RBRVS could be used in this regard (Zuckerman et al., 2014). A related issue is relativity. In the current RBRVS, values for codes are maintained in a relative manner. In other words, each code in the RBRVS 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 that code, and relativity is only applicable within surgical codes. The impact of incorporating RAND's valuation for surgical codes on nonsurgical codes is something that would need to be explored in the future.

Summary

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 model 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 alternative models that reflect major methodological decisions and trade-offs. While the RAND model addresses 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 counterpoint to RUC valuations of services, there are many key limitations to the model. Future work incorporating clinical input and obtaining more data can help to address these limitations and further refine the RAND model.

References

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

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, 2014. 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

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

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.

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.

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.

Zuckerman, S., R. Berenson, K. Merrell, T. Oberlander, N. McCall, R. Lewis, S. Mitchell, and M. Shrestha, *Development of a Model for the Valuation of Work Relative Value Units: Objective Service Time Task Status Report*, 2014. As of November 13, 2014: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeeSched/Downloads/RVUs-Validation-Urban-Interim-Report.pdf