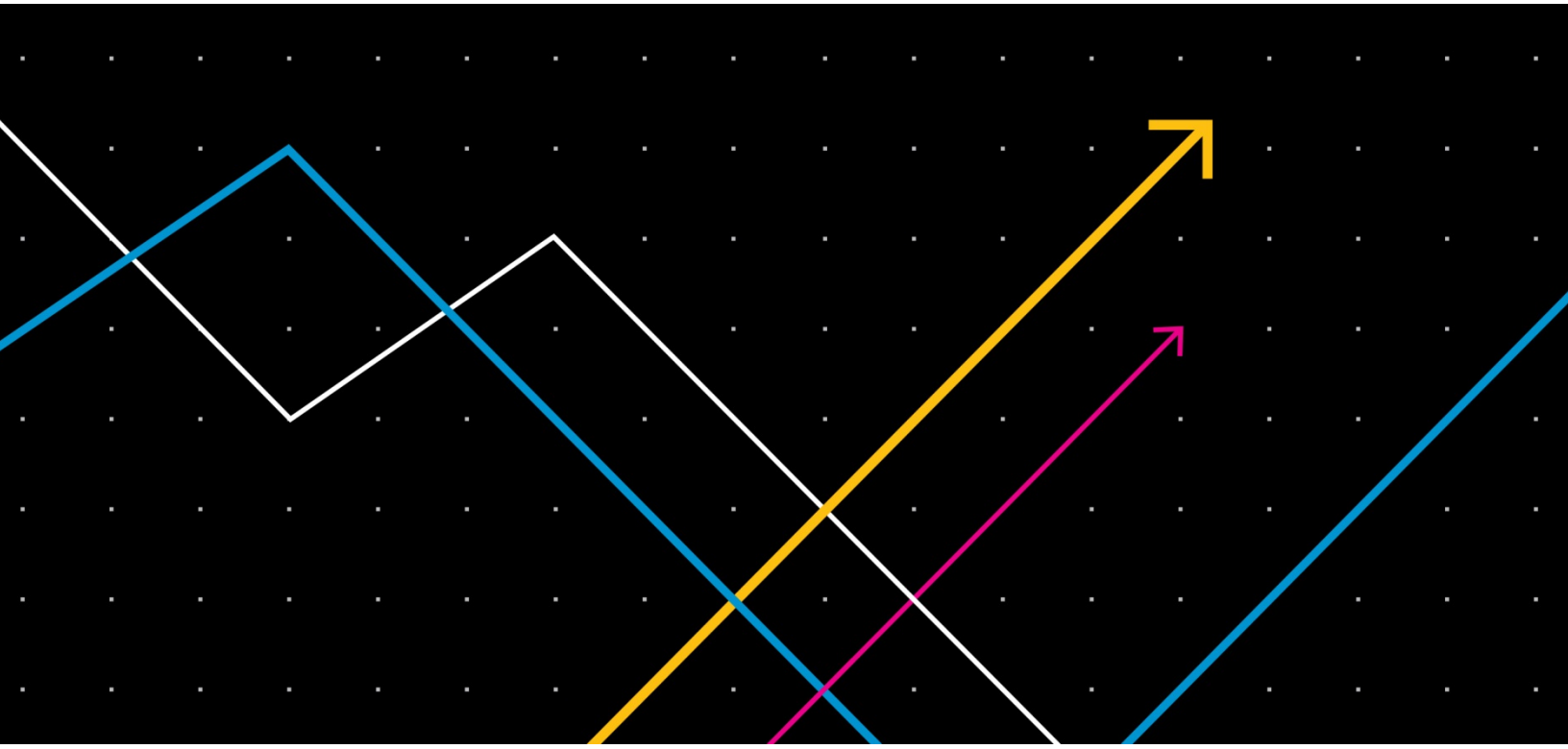


RESEARCH REPORT

Collecting Empirical Physician Time Data

Piloting an Approach for Validating Work Relative Value Units

Stephen Zuckerman, PhD
URBAN INSTITUTE

Katie Merrell, BA
ACTUARIAL RESEARCH CORPORATION (FORMERLY SOCIAL & SCIENTIFIC SYSTEMS, INC.)

Robert Berenson, MD
URBAN INSTITUTE

Susan Mitchell, RHIA
RTI INTERNATIONAL

Divvy Upadhyay, MD, MPH
URBAN INSTITUTE

Rebecca Lewis, MPH
RTI INTERNATIONAL

December 2016

URBAN INSTITUTE

Social & Scientific Systems, Inc.
an employee-owned company

RTI INTERNATIONAL

ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is dedicated to elevating the debate on social and economic policy. For nearly five decades, Urban scholars have conducted research and offered evidence-based solutions that improve lives and strengthen communities across a rapidly urbanizing world. Their objective research helps expand opportunities for all, reduce hardship among the most vulnerable, and strengthen the effectiveness of the public sector.

Contents

Acknowledgments

This report was funded by the Centers for Medicare & Medicaid Services. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at www.urban.org/support.

The authors would like to acknowledge the contributions of many organizations and individuals without whom this study would not have been possible. First and foremost, we would like to thank the staff at the various health systems that provided the empirical time data for our analysis and the physicians who served as clinical reviewers. We also benefited greatly from input and guidance provided by staff at the Centers for Medicare & Medicaid Services, including Kathy Bryant, Chava Sheffield, Ryan Howe, Christina Ritter, Donald Thompson, and Sara Vitolo. Nancy McCall (formerly of RTI) and Peter Braun helped in the early stages of this project with data collection design. Finally, additional staff at the Urban Institute (Nicole Lallemand, Emily Hayes, and Jacob Fass), Social & Scientific Systems, Inc. (Kanul Amrahova and Tyler Oberlander), and RTI (Madhu Shrestha) helped at various stages of the project.

Executive Summary

This pilot project is part of the Centers for Medicare & Medicaid Services' (CMS) efforts to address potentially misvalued services in the Medicare Physician Fee Schedule (PFS), as identified in the Affordable Care Act. The pilot aimed to develop a validation process for the work relative value units (RVUs) used in the fee schedule for both new and existing services to establish payment for the work of physicians or nonphysician practitioners. The project focused on the physician service times used in establishing physician work RVUs. There were essentially two distinct elements of the project: developing empirical measures of physician service times and considering the implications of these estimates for physician work RVUs.

The bulk of the project was devoted to the development of empirical time estimates based on data from several health systems with multispecialty group practices. We collected two types of data for 60 services defined by the Healthcare Common Procedure Coding System (HCPCS): (1) administrative data from electronic health records (EHRs) and (2) direct observation data, for which project or practice staff observed and documented the time needed to provide specific services to individual patients. Based on our analysis of the data, we drew the following conclusions:

- The empirical time data for the 60 HCPCS codes suggest that there may be systematic overvaluations of times for these services within the PFS and, by implication, undervaluation of other services.
- The potential RVU distortions that we detected may also be related to some unrepresentative vignettes and inaccuracies in the tasks outlined in the service descriptions used by the American Medical Association/Specialty Society Relative Value Update Committee (RUC).
- A broader study designed to collect empirical time data that can be used to validate work RVUs for more services seems feasible if health systems give it high priority and adequate resources are available.

As documented in our June 2014 Objective Service Time Task Status Report, we encountered a number of unexpected challenges in recruiting and retaining sites for data collection and in collecting both EHR and direct observation data (Zuckerman et al. 2014). We eventually succeeded in collecting empirical time data through direct observation at three sites. These three sites were located in three distinct regions of the country: New England, Middle Atlantic, and Pacific. Two of the sites were also able to provide data from EHRs.

To augment and assess data from these sites, we conducted clinical expert interviews with 30 physicians, soliciting their time estimates for specific studied services and discussing how their estimates compared to both the current PFS time values and the empirical time data we collected. These 30 physicians were drawn from six specialties and reviewed information for 25 services.

We collected reasonably reliable and accurate empirical time estimates for 60 HCPCS codes. This type of work RVU validation information had not been publicly available prior to our study. Overall, our results suggest that current PFS physician intraservice time values tend to be higher than the empirical data we collected through direct observation and EHRs for the services we studied. For the 60 HCPCS codes in this study, our empirical time values were more than 10 percent lower than the PFS time for 42 services, but more than 10 percent higher for 8 services; our values were roughly equal to PFS time (within 90 percent to 110 percent of one another) for 10 services. The extent to which PFS intraservice times differ from empirical intraservice times varies considerably across service types. Table 3 shows that the ratio of PFS intraservice time to the median empirical intraservice time was greater than 2 for imaging and test interpretations, but close to 1 for inpatient procedures with global periods.

Section 1848(c)(2)(C)(i) of the Social Security Act requires that physician work RVUs reflect both service time and service intensity. If either is inaccurate, then relative work values are distorted. We examined the implications of our revised service time estimates on the implied intensity of studied services, assuming current intraservice work RVUs are correct. Since our time estimates are typically lower than current service times for some types of services, we found that service intensities based on these new time data are often substantially higher than intensities based on PFS times. This suggests that current intraservice work RVUs are overstated.

Staff at all three sites reported that the descriptions of physician tasks that underlie physician time and work values often do not conform to the actual tasks performed by physicians in their practices. This issue, also described in the Status Report (Zuckerman et al. 2014), has remained a focus of the project, since it directly affects how people estimate the time and work associated with specific HCPCS codes. For the clinical expert review of the time data we collected, we used the same descriptions used in the current valuation process. In addition to soliciting time estimates, we asked the physicians we interviewed to comment on the service descriptions and to indicate whether they perceived the specific vignette associated with each HCPCS code to describe a "typical" patient.

The clinical expert review of the vignettes, service descriptions, and time estimates produced several important findings. The RUC, in the process of recommending Current Procedural Terminology (CPT) code valuations to CMS, creates vignettes describing typical clinical scenarios for each service.

Our reviewers raised concerns that some of the vignettes described clinical situations that were not typical and thus may lead to biased PFS time estimates, usually but not always biased toward more time. Similarly, the reviewers felt that some elements of the preservice, intraservice, and postservice descriptions did not accurately reflect current clinical practice, potentially skewing the results of the RUC surveys that use them and, by extension, RUC's advice to CMS. In addition, the reviewers' intraservice time estimates typically fell between study times and PFS times. Reviewers identified a number of factors that may affect the accuracy of service time and work data, including the use of new technology and the availability of clinical staff to provide a portion of the activities associated with the service. These nuances could contribute to distortions in the valuation of PFS work RVUs and point to the need for regular review and updating of the vignettes and service descriptions.

Despite the challenges we faced in the course of our research, this pilot study demonstrates the feasibility of collecting and processing empirical time data and offers important insights that can guide future validation of work RVUs for physician services. First, direct observation efforts should first log the activities of a set of physicians within a specific time period and then identify the services they provide, rather than starting with a set of targeted services. Second, since health systems often require their own personnel to do direct observation, a formal mechanism should be in place to assure quality control. Third, pre- and postservice tasks need to be assessed, even though they are more difficult to observe than intraservice tasks; our clinical experts raised serious questions about whether descriptions of pre- and postservice activities reflect typical practice. Fourth, working with data from EHRs is difficult because EHR systems vary and are not consistently employed across departments within a single health system. Fifth, EHR data are likely to be available only for certain types of services and not for others (e.g., office-based procedures or tests), and direct observation is more feasible for some types of services than for others. Finally, practices that do not rely on HCPCS codes for fee-for-service billing do not track HCPCS codes particularly well for all services and probably should not be used in studies similar to this one.

The findings of this pilot study suggest that the current approach to estimating time and work results in differentially inflated time and work values throughout the PFS, causing inconsistently inaccurate payment rates for physician services. We suggest that CMS shift from its current approach, which relies on specialty society surveys and the RUC to estimate time and work, to empirical determination of time for the most common, high-dollar-volume services. But however CMS chooses to proceed, our research indicates a critical need for improvement in the timeliness and consistency of time and work valuation.

Background

This pilot project is part of the Centers for Medicare & Medicaid Services' (CMS) efforts to address potentially misvalued services in the Medicare Physician Fee Schedule.[1] The broader effort aims to develop a validation process for the work relative value units (RVUs) used in the fee schedule for both new and existing services. It is designed to provide CMS with a process for reviewing proposed work RVUs, assessing how reasonable they are relative to external data and assuring that the relativities within the fee schedule are internally consistent within families of services as well as across families.

Work RVUs reflect both the time it takes the clinician to provide a service and the intensity of the service. With intensity reflecting factors such as technical skill, physical effort, mental effort and judgment, and stress due to patient risk, time is the component of the work RVU most amenable to empirical measurement. service time estimates are currently based on surveys conducted by specialty societies for the American Medical Association/Specialty Society Relative Value Update Committee (RUC). These surveys present clinical reviewers[2] with vignettes that describe a "typical" patient presenting for each service as well as a list of the specific activities that physicians perform during each of three service periods: preservice, intraservice, and postservice. These vignettes and service elements describe the context in which respondents are asked to provide pre-, intra-, and postservice time estimates and, ultimately, total work values.

This project is focused on developing new estimates of intraservice time as a central element in our support of CMS efforts to validate the work RVUs for a selected set of services. The project has three key elements:

1. Obtain empirical time estimates for a group of services from several physician practices or health care systems
2. Compare these empirical time estimates with current fee schedule time data and assess the implications of these data for physician work values
3. Review empirical time estimates and service descriptions with a series of clinical experts in relevant specialties

Following the model of previous studies (summarized below), we collected empirical time data by acquiring administrative data for some types of services and conducting direct observation for other types. We used these data sets to develop time estimates for each of the services selected for study, compared the estimates to existing time values and implied service intensity, and finally submitted both of these to clinical experts for assessment. We described the initial stages of the data collection process

in our interim report, focusing on the unexpected challenges we faced in recruiting and retaining sites for data collection and in collecting both types of data (Zuckerman et al. 2014). In this final report, we provide a brief summary of key elements of the interim report before we analyze the new evidence.

This report presents the details of our empirical service time analyses, including:

- Discussion of our approach to collecting empirical time data
- Recruitment of and interviews with physician experts as part of the clinical review process
- Analysis of the time data in comparison to current PFS values
- Assessment of the analytic findings, including the input of the physician experts
- Identification of challenges that arose in the process of data collection

The following sections address each topic in turn, describing the approaches we used in light of the issues and challenges that arose during this pilot. We largely overcame the data collection challenges documented in our interim report, but these difficulties limited the number of sites and the number of HCPCS codes we were able to study. However, none of these data collection challenges forced us to stray from our original goals of collecting empirical time data, subjecting them to clinical review, and assessing how these new data could shape the validation of PFS work RVUs.

Collecting Empirical Service Time Data

The goal of this project is to develop independent, empirical measures of service-level times for PFS services. Prior research has shown that some of the Medicare fee schedule's estimates of service time, based upon surveys of physicians by specialty societies, are considerably higher than estimates obtained from other data sources, such as operating room logs (McCall, Cromwell, and Braun 2006). That is, the survey times on which both work and practice expense RVUs are based may diverge from empirical measures of service time collected through administrative systems or direct observation.

These differences in service time between the PFS and other sources may stem from changes that have occurred since the PFS was introduced almost 25 years ago. Through the substitution of new technologies, such as the picture archiving and communication system (PACS) for imaging, the time for a particular service may be much less than what the fee schedule work RVUs suggest. Many of these services have never been restudied or validated, except by the RUC. RUC review is limited by its use of time estimates obtained from surveys of physicians conducted by specialty societies as part of their requests for changes in work RVUs.

Because of concerns about the accuracy of time estimates in the fee schedule, the Medicare Payment Advisory Commission (MedPAC) funded a study to assess the feasibility of using empirical time data to establish or update work RVUs (Braun and McCall 2011). The study team conducted key informant interviews with five organizations representing a broad cross section of sites. Although none of the interviewed organizations had collected clinical service times linked to specific HCPCS codes, they expressed limited concern about their ability to link clinical service times to HCPCS codes, regardless of data collection method. Furthermore, the MedPAC study concluded that intraservice clinical time was best captured for major surgical procedures in EHR data, but pre- and postservice times were not well captured. The interviewed organizations doubted that their electronic data systems captured time for ambulatory surgical centers or other types of procedures such as endoscopy, radiology, and cardiac catheterization. None of the organizations believed they captured clinical service time for office-based procedures such as tests or skin lesion removal.

Interviewees noted that the potential to use EHR data systems may exceed their organization's use to date and warranted further evaluation. However, the organizations expressed a great deal of uncertainty about the capability of their electronic systems to capture clinical service time, and many

noted that not all elements of their electronic data systems had been fully implemented. The organizations had varying degrees of sophistication and experience with the collection of clinical service time through direct observation. Although no organization had collected time for the purpose of payment under the fee schedule, all organizations felt that direct observation was feasible given their prior experience with collecting clinical service time for other purposes.

Our project built on the MedPAC study by collecting empirical time data from both administrative electronic data and direct observation. We worked with study sites to determine the types of administrative data available and to develop a plan for direct observation.

Our research team worked closely with CMS to identify about 100 HCPCS codes for the study. We considered three factors when identifying HCPCS codes to study:

- Did the Affordable Care Act include the services among those that are thought to be at risk for being misvalued?[3]
- Were the services important to Medicare, either because of their total spending or for other policy reasons (including having global service periods or serving as one of the multiple points of comparison)?
- Would this mix of services allow us to test methods in a variety of clinical practice settings, yet still be limited to a sufficiently small number of specialties to make the clinical expert review of the study findings manageable?

Balancing these considerations, we developed a list of 117 HCPCS codes for the study, as shown in appendix A. They included 29 services from the then current CMS list of potentially misvalued services, 71 procedures with global periods of varying lengths, and 12 services that have been used as multiple points of comparison in the development and refinement of work relative values.

We worked with CMS staff to develop a list of potential sites for data collection. In developing this list, we considered site administrators' interest in participation, health IT capabilities, experience with direct observation, and the site's mix and volume of clinical services provided. The process of identifying, recruiting, assessing, and engaging sites was much more complicated than originally anticipated; we were only able to engage three sites, after having approached nearly twenty potential sites. Many of the practices that declined expressed interest in participating in a study like this but faced logistical or organizational barriers that made participation too difficult. The three sites were located in three different regions of the country: New England, Middle Atlantic, and Pacific. All three sites conducted direct observation, but only two provided data from EHRs. We recognize that these three

sites were very much a sample of convenience and should not necessarily be viewed as representative of other health systems.

To document the time associated with intraservice work, we used the preservice, intraservice, and postservice descriptions provided by the RUC. Using these descriptions created some difficulties that may have affected our measures. First, the descriptions vary in detail about the specific tasks included in the service. Some services have vague descriptions, while others are quite detailed. Consequently, observers may not consistently attribute tasks to the pre-, intra-, or postservice period. The descriptions are also inconsistent about whether or not tasks performed by nonphysician clinical staff are included. In light of these inconsistencies, we developed a data collection tool that allowed observers to indicate which elements from the service descriptions were performed and by whom. The data collection tool also allowed observers to indicate additional tasks that were performed as part of the service, beyond those included in the service descriptions. This approach allowed us to examine the specific service elements provided and who provided them, which could help CMS understand how closely the service descriptions reflect current service provision. Our interviews with clinical experts about the service descriptions helped explain why discrepancies may exist between the time data we collected and the PFS time estimates.

Analytic Approach and Methods

This section describes our approach to the two analyses we conducted. The first part reviews how we prepared the empirical time data for analysis and determined which cases would be included in the analysis. The second part explains how we developed a clinical expert review process and what types of information we solicited from the participants.

Preparing the Empirical Time Data for Analysis

The project team collected time data using two sources: direct observation and EHRs. Preparing these data for analysis required clear definitions of the analysis variable (in this case, intraservice time) and the cases suitable for inclusion in the analysis.

Empirical Measure of Intraservice Physician Time

Data from direct observation and EHRs were used to develop time measures as comparable as possible to the service-level time values currently used in the PFS. This process involved a number of decisions and calculations.

We had originally planned to analyze the relationship between direct observation- and EHR-derived time estimates for the same patient cases, in order to assess how closely the two time estimates matched. If there appeared to be discrepancies between the two data collection approaches, we planned to make some adjustments. However, because we were only able to acquire both types of data for a few patients, we revised our approach. Our new approach combined the direct observation- and EHR-derived estimates into a single empirical time measure and treated either source as equally valid. While EHR data may not capture interruptions as well as direct observation data, interruptions are less likely to occur during the intraservice period than during the pre- and postservice periods, particularly in the context of a dedicated operating room or procedure suite. We suspect that any systematic biases in either or both of these measures are likely to be lower for intraservice time than for pre- or postservice time, such that EHR and direct observation data are likely to be the most comparable for intraservice time.

EHR Measure of Intraservice Time. To acquire EHR data, we used a different approach for each of the two sites and an interdepartmental approach for one site. One site used a single major EHR system for all departments for which it provided data, while the other used different products for different departments. This site employed different EHR systems in cardiac catheterization/electrophysiology, gastroenterology, and surgery. Both study sites were asked to provide Excel files containing the following data elements for each patient with a targeted HCPCS code during the defined data collection time period:

- Patient's unique study ID
- Date of service
- HCPCS code(s) for service, based upon the HCPCS billing code assigned by professional coding staff
- Each event recorded during service (e.g., "Pause - Time Out Complete," "Procedure Start/Incision")
- Time stamp for each recorded event

We asked the sites to provide cases that included one of this study's HCPCS codes in their final code assignments. Some cases were erroneously included on the basis of the scheduled service (which included one of this study's HCPCS codes) rather than on the final HCPCS codes assigned by the site coders (which did not include one of this study's HCPCS codes). The project team worked closely with site staff to refine the case selection process during the designated period to capture as many eligible cases as possible. Data were collected for a six-month period at one site and for a twelve-month period at the other. EHR data files typically reported all events with a time stamp in the format *hh:mm:ss*. We calculated the intraservice time in minutes from the start and end time stamps, excluding the minutes associated with any documented interruptions or pauses.

Project staff worked with site staff to map available data from each EHR system to the intraservice period, as shown in table 1. It is not possible to know how closely these start and end times correspond to the intraservice period described by the service description or to that described by direct observation of services. Such systematic discrepancy would affect our assessment of current PFS intraservice time.

TABLE 1

System Times Used to Define Intraservice Period in EHR Data

	Event equivalent to intraservice start time	Event equivalent to intraservice end time
Site 1		
All procedures	Procedure Start/Incision	Procedure Finish/Close
Site 2		
Cardiac catheterization/ electrophysiology laboratory	Case Start	Case End
Endoscopy	Upper Scope In	Upper Scope Out
Colonoscopy	Lower Scope In	Lower Scope Out
Surgery	Surgical Incision Time	Surgery Stop Time

Direct Observation of Intraservice Time. The direct observation process (described in detail in our interim report, Zuckerman et al. 2014) was designed around the notion that the research team would work with participating practices to observe and time study services. Each direct observation event started with an expected HCPCS code for the service to be performed, based on practice scheduling and other information. The data collection tool required observers (1) to record all tasks according to the service period (pre-, intra-, or post-) to which they pertained, based on the service descriptions; and (2) to indicate who was performing each task, in order to separate time spent by the physician from that spent by other members of the care team. Sites also recorded the time associated with any interruptions that occurred, allowing us to calculate the net service time of interruptions.

The specific service descriptions that guided data collection were based on the preservice, intraservice, and postservice descriptions used by the RUC in their time and work surveys. Site staff and clinicians remarked that the RUC list did not conform closely to their clinical processes for many study services. For this project, observers were trained to record whatever listed tasks were performed and were provided space to record any additional tasks. Putting aside the larger question of whether the service descriptions consistently include tasks in the phase of service when they typically occur, our option to add additional tasks introduced the possibility of miscategorization; for example, an observer might add an "intraservice" task that should have been labeled pre- or postservice. To reduce this risk, we advised observers to use the first and last tasks listed in the intraservice period as the anchors to determine where to assign time for unlisted tasks. The observed mismatch between listed tasks and actual clinical practice raises important questions about the extent of the discordance and its effect on physician responses to RUC surveys. In light of these concerns, we evaluated the accuracy of the service descriptions in the clinical expert review.

CASES INCLUDED IN ANALYSIS

We could not include all available direct observation and EHR data in the analysis for several reasons. First, not all cases included one of the study's 117 HCPCS codes among their final assigned codes. EHR data were selected based on the final HCPCS codes assigned to each case, so only cases with one of the study's HCPCS codes were pulled from the EHR data. However, direct observation cases might have an ultimate code that was *not* one of the study's HCPCS codes, even though we targeted the study's HCPCS codes for observation. A number of observed cases did not end up with a final HCPCS code among those selected for our study and were thus excluded from our analysis. We also excluded many direct observation cases for which study sites did not provide final HCPCS codes.

In both EHR and direct observation data, individual cases were often assigned multiple HCPCS codes. Notably, over 25 percent of the EHR cases we received had more than one assigned HCPCS code. We were unable to include these cases in our analysis, since it was not possible to allocate measured time to each of the individual codes in the record.

However, we made one exception to this rule for cases assigned the coronary artery bypass graft (CABG) code, for which an add-on code was also used. Generally, CABG procedures are coded with a base HCPCS code that reflects a single-vessel procedure and, as appropriate, an add-on HCPCS code when more than one vessel is involved in the surgery. HCPCS code 33533 represents a single arterial CABG and can appear alone in the case of a single-vessel procedure. It may also appear on a record with one of two HCPCS add-on codes (33518 for two vessels, or 33519 for three vessels) that cannot appear alone on a record. We included records with multiple final HCPCS codes when one code was 33533 and the other code was either 33518 or 33519. All three of these codes were included in the initial 117 HCPCS codes selected for this study. For this analysis, we combined the PFS time and work RVU data for the two HCPCS codes (33533 + 33518; 33533 + 33219) for comparison with the objective time measure. All other cases had only one HCPCS code.

In sum, cases were included in the analysis if all of these criteria were met:

- A final HCPCS code is available for the observed event.
- Only one HCPCS code is assigned to the event (unless the additional codes are 33518 and 33519, which are included when occurring with 33533 and no additional codes).
- The final assigned code is one of 117 previously identified HCPCS codes.
- *For direct observation cases only:* The case includes tasks associated with the intraservice period.

- *For direct observation cases only:* The case includes intraservice tasks performed by a physician or a nonphysician practitioner.[4]

After applying these rules to the raw data, we omitted HCPCS codes that had fewer than eight cases. Although eight is a small and somewhat arbitrary number, we chose this cutoff because it reflected the service volumes we encountered during direct observation. Service volume was lower at each study site than we expected based on the Medicare volume data that guided our site selection process. We applied the same eight-case cutoff to the EHR data for consistency and because EHR samples were not uniformly higher than those from direct observation. Of the initial selection of 117 HCPCS codes (appendix A), 60 codes were included in our final analysis. The median of intraservice time from direct observation and EHR data is the key HCPCS-level analysis variable and is based on a total of 7,405 cases with empirical time data for 60 HCPCS codes.

GROUPING SERVICES FOR ANALYSIS

We grouped codes into four types of services: tests (imaging and other), physician's office, outpatient department or ambulatory surgery center (OPD/ASC), and inpatient hospital. The research team assigned services to the "tests" group based on their clinical knowledge and judgment. Remaining services were assigned to the other three groups by their dominant service location, as captured by Medicare claims data reported in the RUC database.[5]

Within each service type, we grouped codes into families of closely related services. Codes within a service family are typically listed under the same heading in the CPT taxonomy, with the first three or four digits of the HCPCS code common to all services in the family. For example, our study codes include three variations of cystoscopy and three variations of thigh fracture surgeries. In three cases—intestinal surgery, CT/MRI, and noninvasive cardiac testing—we developed broader families, composed of services with only the first two digits in the HCPCS taxonomy and a core activity or technology in common. Nineteen codes that are not closely related to other study codes were not placed into a family but were still classified into one of the four types of services. These services are not included in the small number of family-based analyses presented below.

Clinical Expert Review Methods

We conducted semistructured interviews with physicians to assess the face validity of the empirical time estimates for intraservice elements obtained through direct observation and EHRs. For a subset of

services, we also assessed the representativeness of HCPCS code-specific vignettes and the accuracy of the service descriptions of pre-, intra-, and postservice elements used in the RUC's specialty society survey process. Though we focused on time in this study, we also wanted to understand how well the vignettes and service descriptions described work activities and how the activities themselves may have changed over the years because of technological innovation, personnel substitution, changes in clinical practice, or other variables.

Services. We selected a subset of 25 HCPCS codes to include in the clinical expert review process. Most of these were drawn from the 60 HCPCS codes for which we had empirical time data, but we added several other HCPCS codes from the original pool of 117 to better represent the activities of the specialties we included. When selecting HCPCS codes for inclusion in this part of the study, we attempted to include codes representing a range of intraservice times consistent with the typical work of the specialty. For example, for urology, we included both office-based and inpatient procedures, since this specialty provides both types of services. We organized the codes into six specialty groups, shown in table 2, based on the primary specialty that provides the service.

In response to the feedback we received from clinic site staff and the discrepancies we observed between PFS and empirical time data, we decided to study the vignettes and the preservice, intraservice, and postservice descriptions as part of the clinical review process. We found that it was hard to assess the empirical time data we collected without having a clear picture of what was involved in those services. Thus, we assessed vignette representativeness and service description accuracy for all four types of services: inpatient procedures, outpatient/ambulatory surgical center procedures, office-based procedures, and imaging/test interpretations.

TABLE 2

Services Included in the Clinical Expert Review

Specialty	HCPCS code	Code Descriptor
Noninvasive cardiology	93010	Electrocardiogram, routine ECG with at least 12 leads; interpretation and report only
Noninvasive cardiology	93015	Cardiovascular stress test using maximal or submaximal treadmill or bicycle exercise, continuous electrocardiographic monitoring, and/or pharmacological stress; with supervision, interpretation and report
Noninvasive cardiology	93306	Echocardiography, transthoracic, real-time with image documentation (2D), includes M-mode recording, when performed, complete, with spectral Doppler echocardiography, and with color flow Doppler echocardiography
Orthopedics	27130	Arthroplasty, acetabular and proximal femoral prosthetic replacement (total hip arthroplasty), with or without autograft or allograft
Orthopedics	27236	Open treatment of femoral fracture, proximal end, neck, internal fixation or prosthetic replacement

Specialty	HCPCS code	Code Descriptor
Orthopedics	27447	Arthroplasty, knee, condyle and plateau; medial AND lateral compartments with or without patella resurfacing (total knee arthroplasty)
Orthopedics	29881	Arthroscopy, knee, surgical; with meniscectomy (medial OR lateral, including any meniscal shaving) including debridement/shaving of articular cartilage (chondroplasty), same or separate compartment(s), when performed
Ophthalmology	66984	Extracapsular cataract removal with insertion of intraocular lens prosthesis (1 stage procedure), manual or mechanical technique (e.g., irrigation and aspiration or phacoemulsification)
Ophthalmology	67028	Intravitreal injection of a pharmacologic agent (separate procedure)
Ophthalmology	92133	Scanning computerized ophthalmic diagnostic imaging, posterior segment, with interpretation and report, unilateral or bilateral; optic nerve
Ophthalmology	92134	Scanning computerized ophthalmic diagnostic imaging, posterior segment, with interpretation and report, unilateral or bilateral; retina
Gastroenterology	43235	Esophagogastroduodenoscopy, flexible, transoral; diagnostic, including collection of specimen(s) by brushing or washing, when performed (separate procedure)
Gastroenterology	43239	Esophagogastroduodenoscopy, flexible, transoral; with biopsy, single or multiple
Gastroenterology	45378	Colonoscopy, flexible; diagnostic, including collection of specimen(s) by brushing or washing, when performed (separate procedure)
Gastroenterology	45380	Colonoscopy, flexible; with biopsy, single or multiple
Gastroenterology	45385	Colonoscopy, flexible; with removal of tumor(s), polyp(s), or other lesion(s) by snare technique
Radiology	70553	Magnetic resonance (e.g., proton) imaging, brain (including brain stem); without contrast material, followed by contrast material(s) and further sequences
Radiology	71020	Radiologic examination, chest, 2 views, frontal and lateral
Radiology	71250	Computed tomography, thorax; without contrast material
Radiology	72148	Magnetic resonance (e.g., proton) imaging, spinal canal and contents, lumbar; without contrast material
Urology	52000	Cystourethroscopy (separate procedure)
Urology	52224	Cystourethroscopy, with fulguration (including cryosurgery or laser surgery) or treatment of MINOR (less than 0.5 cm) lesion(s) with or without biopsy
Urology	52601	Transurethral electrosurgical resection of prostate, including control of postoperative bleeding, complete (vasectomy, meatotomy, cystourethroscopy, urethral calibration and/or dilation, and internal urethrotomy are included)
Urology	55700	Biopsy, prostate; needle or punch, single or multiple, any approach
Urology	55866	Laparoscopy, surgical prostatectomy, retropubic radical, including nerve sparing, includes robotic assistance, when performed

Source: AMA/Specialty Society RVS Update Committee database.

Of the 25 HCPCS codes we selected, six were not included in the analysis of empirical time estimates because they did not have sufficiently large numbers of observations. These HCPCS codes were included in the clinical review so that we could study the category of imaging/test interpretations and explore several procedural services within specialties, from office-based procedures to inpatient surgeries.[6] Despite the absence of empirical data, the comparison to PFS intraservice time estimates provided useful insight about time and service description accuracy.

We assessed vignette representativeness and intraservice service description accuracy for all 25 HCPCS codes. We also reviewed the pre- and postservice descriptions for two distinct codes in each specialty (except in ophthalmology, for which we reviewed all four codes). We specifically discussed the following factors:

- Whether vignette selection is important in making reliable estimates of physician time and work
- Whether vignettes for the selected codes in each specialty are typical and, if not, how that might affect time and work
- How accurately service descriptions reflect clinical practice—specifically, whether the descriptions reflect technological advances, personnel substitution, and other factors that might lead to changes in the nature of physician activities

Other sources of time and work variations based on the physician's practice experience were explored as well.

Physician Interviews

We conducted interviews with five physicians in each of the six specialties included in the clinical review. These discussions allowed us to explore the reasons for variation in time and work estimates, vignette representativeness, and service description accuracy.

With only five interviewees per specialty, we used a convenience sample, rather than a representative sample, of physicians across the country. We worked with five multispecialty group practices (one a source of empirical time data) located in the four census regions of the country (East, Midwest, South, and West). We coordinated with the groups' chief executive officers and chief medical officers to identify physicians within the six specialties willing to participate in semistructured interviews lasting between 30 and 45 minutes. We then approached each physician individually to obtain consent and to schedule the interview, which took place in January 2016. The physicians were sent PFS code-specific vignettes and intraservice descriptions for the codes in their specialty as well as the pre-and postservice descriptions for the two HCPCS codes to be discussed in detail. The project's clinical director used input from clinical reviewers to determine when inconsistencies in vignettes or service descriptions were significant enough to potentially affect estimates of time.

After discussing the vignettes and service descriptions, we closed the interviews by asking physicians to estimate what they considered to be typical intraservice times, based on what they consider to be actual intraservice physician activities, whether or not their assessments matched RUC service descriptions. We encouraged each respondent to provide a point estimate, but some were more comfortable providing a time range. When physicians provided time ranges, we converted them to the middle point of the ranges for comparison.

The interviews were confidential and received Institutional Review Board (IRB) approval from the Urban Institute. The clinical director of the project oversaw a team of four interviewers including himself, the project's principal investigator, and two physician researchers. We transcribed all of the physician interviews and produced summaries of the transcripts for each of the six specialties (see appendix B). These summaries were used in the analysis below.

Results

Empirical Time Analysis

This section presents an analysis of our HCPCS-level measure of intraservice time and its potential implications for physician work values. Most of the 60 HCPCS codes for which we had data are closely related to at least one other code included in the study, so we created 14 service families for 46 HCPCS codes to identify whether the relationship between PFS and study time was consistent within service families. A family consists of codes found in the same section of the Common Procedural Terminology code manual; physician tasks for these codes are similar. These small service families, which fit into four broader types of service, are used in some analyses to discern patterns among and across groups of codes. The four types of services are (1) office-based procedures; (2) outpatient department (OPD) or ambulatory surgery center (ASC) procedures, with or without global period; (3) inpatient procedures with global periods; and (4) imaging and other test interpretations. (Fourteen of the study's HCPCS codes for which we have empirical data do not relate to another code and thus were not assigned to a service family. However, these codes do fit into our types of service and are included in analyses that do not rely on service family.[7]) As described above, we have combined the direct observation and EHR data into a single intraservice time analysis variable, measured as the median intraservice time for all study cases with that HCPCS code. The potential implications of this approach are considered at the end of this section.

Empirical Time Estimates for Selected Services

Overall, our median intraservice time values tended to be lower than current PFS values.[8] Over three-quarters of the study's 60 HCPCS codes have a median value of study intraservice time below the current PFS value, as shown in table 3. Of the study's 60 HCPCS codes, 42 have study times more than 10 percent lower than the current PFS value (table 4), 10 have study times roughly equal to the PFS time (i.e., within 10 percent), and 8 have study times more than 10 percent higher than the PFS time.

TABLE 3

Comparison of PFS and Empirical Intraservice Physician Time, by HCPCS Code, Type of Service, and Service Family

Physician office-based procedures (procedures substantially performed in an office setting)

HCPCS code	Service family	Brief service descriptor	2016 PFS intraservice time (min)	Median empirical intraservice time (min)	N	Ratio of PFS to median empirical intraservice time
52000	Cystoscopy (urinary bladder)	Cystoscopy	15	14	54	1.07
52224	Cystoscopy (urinary bladder)	Cystoscopy and treatment	30	19	45	1.58
52281	Cystoscopy (urinary bladder)	Cystoscopy and treatment	20	16	36	1.25
11042	No family	Deb subq tissue 20 sq cm/<	15	25	37	0.60
17110	No family	Destruct b9 lesion 1–14	7	15	16	0.47
55700	No family	Biopsy of prostate	15	13	30	1.15

Outpatient department/ambulatory surgical center procedures, with or without global period (procedures substantially performed in an ambulatory surgical center or outpatient hospital setting)

HCPCS code	Service family	Brief service descriptor	2016 PFS intraservice time (min)	Median empirical intraservice time (min)	N	Ratio of PFS to median empirical intraservice time
43235	GI endoscopy	Egd diagnostic brush wash	15	5	61	3.07
43239	GI endoscopy	Egd biopsy single/multiple	15	6	227	2.34
45378	Colonoscopy	Diagnostic colonoscopy	25	20	304	1.28
45380	Colonoscopy	Colonoscopy and biopsy	28	21	333	1.31
45385	Colonoscopy	Colonoscopy w/lesion removal	30	22	120	1.39
G0105	Colonoscopy	Colorectal scrn; hi risk ind	25	18	29	1.39
47562	Laparoscopic removal of gall bladder	Laparoscopic cholecystectomy	80	66	359	1.21
47563	Laparoscopic removal of gall bladder	Laparo cholecystectomy/graph	90	82	213	1.10
66982	Cataract	Cataract surgery complex	33	22	161	1.50
66984	Cataract	Cataract surg w/iol 1 stage	21	18	1565	1.17
93458	Cardiac angiography	L hrt artery/ventricle angio	45	22	86	2.05
93459	Cardiac angiography	L hrt art/grft angio	50	43	9	1.16
23412	No family	Repair rotator cuff chronic	100	78	140	1.28
29827	No family	Arthroscop rotator cuff repr	120	81	76	1.48
33249	No family	Insj/rplcmt defib w/lead(s)	120	42	29	2.86
49505	No family	Prp i/hern init reduc >5 yr	70	60	353	1.17
50590	No family	Fragmenting of kidney stone	60	35	260	1.71
52601	No family	Prostatectomy (TURP)	75	56	142	1.35
92928	No family	Prq card stent w/angio 1 vsl	76	62	12	1.24

Inpatient procedures with global period

HCPCS code	Service family	Brief service descriptor	2016 PFS intraservice time (min)	Median empirical intraservice time (min)	N	Ratio of PFS to median empirical intraservice time
27130	Hip, knee joint replacement surgery	Total hip arthroplasty	100	87	471	1.15
27447	Hip, knee joint replacement surgery	Total knee arthroplasty	100	83	726	1.20
27236	Thigh fracture surgery	Treat thigh fracture	90	81	32	1.12
27244	Thigh fracture surgery	Treat thigh fracture	75	71	16	1.06
27245	Thigh fracture surgery	Treat thigh fracture	80	86	15	0.93
33405	Heart valve replacement surgery	Replacement of aortic valve	197	203	106	0.97
33430	Heart valve replacement surgery	Replacement of mitral valve	232	201	8	1.15
33533	Cardiac bypass surgery	CABG arterial single	158	214	48	0.74
+ 33518	Cardiac bypass surgery	CABG artery-vein two	208	227	57	0.92
+ 33519	Cardiac bypass surgery	CABG artery-vein three	228	252	55	0.90
44120	Bowel removals and resections	Removal of small intestine	134	212	39	0.63
44140	Bowel removals and resections	Partial removal of colon	150	265	19	0.57
44143	Bowel removals and resections	Partial removal of colon	150	203	8	0.74
44145	Bowel removals and resections	Partial removal of colon	180	241	34	0.75
44160	Bowel removals and resections	Removal of colon	120	159	42	0.76
44204	Bowel removals and resections	Laparo partial colectomy	180	142	17	1.27
44205	Bowel removals and resections	Lap colectomy part w/ileum	165	128	11	1.29
23472	No family	Reconstruct shoulder joint	140	112	65	1.25
27134	No family	Revise hip joint replacement	240	132	33	1.82
33208	No family	Insert heart pm atrial & vent	60	46	114	1.30
35301	No family	Rechanneling of artery	120	125	42	0.96
55866	No family	Laparo radical prostatectomy	180	176	95	1.02
63047	No family	Remove spine lamina 1 lmbr	90	50	93	1.80

Imaging and other test interpretations

HCPCS code	Service family	Brief service descriptor	2016 PFS intraservice time (min)	Median empirical intraservice time (min)	N	Ratio of PFS to median empirical intraservice time
70450	CT/MRI	CT head/brain w/o dye	10	5	9	2.00
70551	CT/MRI	MRI brain stem w/o dye	18	8	18	2.25
71250	CT/MRI	CT thorax w/o dye	15	16	11	0.94
72141	CT/MRI	MRI neck spine w/o dye	20	5	8	4.00
93010	Noninvasive cardiac testing	Electrocardiogram report	5	0.1*	411	50.00
93015	Noninvasive cardiac testing	Cardiovascular stress test	20	6	15	3.33
93306	Noninvasive cardiac testing	TTE w/Doppler complete	20	5	20	4.00
G0202	Mammography	Screening mammography digital	5	3	8	1.67
G0206	Mammography	Diagnostic mammography digital	7	5	8	1.56
71020	No family	Chest X-ray 2vw frontal&latl	3	3	8	1.20
77080	No family	DXA bone density axial	5	2	24	2.50
88305	No family	Tissue exam by pathologist	25	2	23	12.50

Source: UI/SSS analysis of primary data and PFS 2016 Final Rule public use files (80 FR 70885).
* The median empirical intraservice time was 6 seconds.

TABLE 4

Distribution of Ratio of PFS to Empirical Intraservice Physician Time

Ratio of PFS to empirical intraservice time	Number of HCPCS codes
Under 0.9	8
0.9 to 1.1	10
Over 1.1	42

Source: UI/SSS analysis of primary data and PFS 2016 Final Rule (80 FR 70885) public use files.

The relationship between PFS time and empirical time appears fairly consistent *within* each of the 14 families of closely related services and somewhat different *across* the families (table 5). In nine families, the constituent HCPCS codes all suggest PFS time is too high, as evidenced by a minimum ratio of PFS to study time greater than 1. However, in one family, cardiac bypass surgery, PFS times were consistently low; the maximum ratio was less than 1 across the three constituent codes. The results are mixed for the other four families (thigh fracture, heart valve, bowel removal/resections, and CT/MRI); their study times are higher than PFS times for some codes and lower for others.

TABLE 5

Comparison of PFS and Empirical Intraservice Physician Time, by Service Family

Ratio of PFS time to median empirical time (at HCPCS level)

Service family	Number of HCPCS codes	Min	Median	Max
Physician office-based procedures (procedures substantially performed in an office setting)				
Cystoscopy (urinary bladder)	3	1.07	1.25	1.58
Outpatient department/ambulatory surgical center procedures, with or without global period (procedures substantially performed in OPD/ASC setting)				
GI endoscopy	2	2.34	2.71	3.07
Colonoscopy	4	1.28	1.35	1.39
Laparoscopic removal of gall bladder	2	1.10	1.15	1.21
Cataract	2	1.17	1.33	1.50
Cardiac angiography	2	1.16	1.60	2.05
Inpatient procedures with global period				
Hip and knee joint replacement surgery	2	1.15	1.18	1.20
Thigh fracture surgery	3	0.93	1.06	1.12
Heart valve replacement surgery	2	0.97	1.06	1.15
Cardiac bypass surgery	3	0.74	0.90	0.92
Bowel removals and resections	7	0.57	0.75	1.29
Imaging and other test interpretations				
CT/MRI	4	0.94	2.13	4.00
Noninvasive cardiac testing	3	3.33	4.00	50.00
Mammography	2	1.56	1.61	1.67

Source: UI/SSS analysis of primary data and PFS 2016 Final Rule (80 FR 70885) public use files.
Note: Fourteen of the 60 study codes are not part of a small service family and thus are not included in this table.

There are also differences in the relationship between the two measures across the four broad types of service shown in table 6. While the median ratio of PFS to study time is greater than 1 for all four categories, it is much higher—greater than 2—for imaging and test interpretation. Services typically provided in outpatient departments and ambulatory surgical centers had the next highest ratio, at 1.35, with the other two categories—office-based procedures and inpatient global surgeries—at lower levels, 1.11 and 1.02 respectively.

TABLE 6

Comparison of PFS and Empirical Intraservice Physician Time, by Type of Service

Ratio of PFS time to median empirical time (at HCPCS level)

Type of service	Number of HCPCS codes	Min	Median	Max
Physician office-based procedures (procedures substantially performed in an office-setting)	6	0.47	1.11	1.58
Outpatient department/ambulatory surgical center procedures, with or without global period (procedures substantially performed in OPD/ASC setting)	19	1.10	1.35	3.07
Inpatient procedures with global period	23	0.57	1.02	1.82
Imaging and other test interpretations	12	0.94	2.38	50.00

Source: UI/SSS analysis of primary data and PFS 2016 Final Rule (80 FR 70885) public use files.

In the process of analyzing our empirical time measure, we became concerned about the potential effect of combining direct observation data with EHR data to derive measures of median times for each HCPCS code. As described earlier, we originally planned to begin with a careful analysis of the relationship between direct observation and EHR values for the same case. However, the challenges of obtaining the final EHR data for direct observation cases were greater than anticipated; we were unable to get information on enough cases to support such an analysis. In the analysis above, data for some services are entirely from direct observation, while data for other services are entirely from EHRs, making direct comparison impossible. For example, all of the empirical time data for the 12 HCPCS codes included in the "imaging and other test interpretations" service type come from direct observation, while all of the data for HCPCS codes in the "inpatient procedures with global period" category come from EHRs.

Our analysis shows that the small service families for which our study data have median intraservice times lower than current PFS times are derived from both direct observation and EHR sources. This suggests that the data source is not driving our results. For example, our study times for the GI endoscopy HCPCS codes and for the CT/MRI interpretation HCPCS codes suggest that the current PFS work RVUs in these two families are over two times too high, yet on average, cases with the GI HCPCS codes derive 99 percent of their data from EHR cases, whereas cases with the CT/MRI HCPCS codes come from direct observation only.

Implications for Physician Work Values

A key objective of this pilot project was to develop a method to validate the work RVUs in the PFS. It focused on the feasibility of developing new estimates of service time, since time is the key driver of the variance in work and many have raised concerns about the validity of some current time values. We examine the potential impact of the new empirical time measure on work RVUs by studying two alternative ratios of work to time (known as intraservice work per unit time, or IWPUT), each based on the same current work RVUs.

Why does the implied intensity matter, and how should it be analyzed? Intensity is a derived measure that is not directly measured. Nonetheless, the concept of implied intraservice intensity is well established in the context of reviewing and establishing physician work values. Underlying the assumption that work RVUs reflect the relative work across services is the notion that if HCPCS codes with the same intraservice time have different intraservice values, this is due specifically to differences in intraservice intensity. Analyzing intraservice intensity among services—essentially, the relationship between intraservice time and intraservice work—can help identify potential problems in valuations.

Since this project only captured intraservice time, we were unable to directly analyze total work and intensity. Instead, we used the building block method to develop an estimate of intraservice work.[9] This method starts by assuming a fixed intensity for pre- and postservice activities across all services. This assumption, along with PFS preservice and postservice time estimates, allows us to calculate the work implied for pre- and postservice activities. In addition, the work associated with the evaluation and management (E&M) services included in the global service periods can be calculated simply as the sum of the work values for all E&M services assumed to be provided during the global period. Removing the pre- and postservice work and the global E&M work from total physician work for the service leaves an implied intraservice work value. In turn, the implied intraservice intensity is the ratio of this implied intraservice work value to intraservice time. Using this implied intraservice work value, we calculated the implied intraservice intensity using current PFS intraservice time and, alternatively, using our empirical intraservice time estimates.

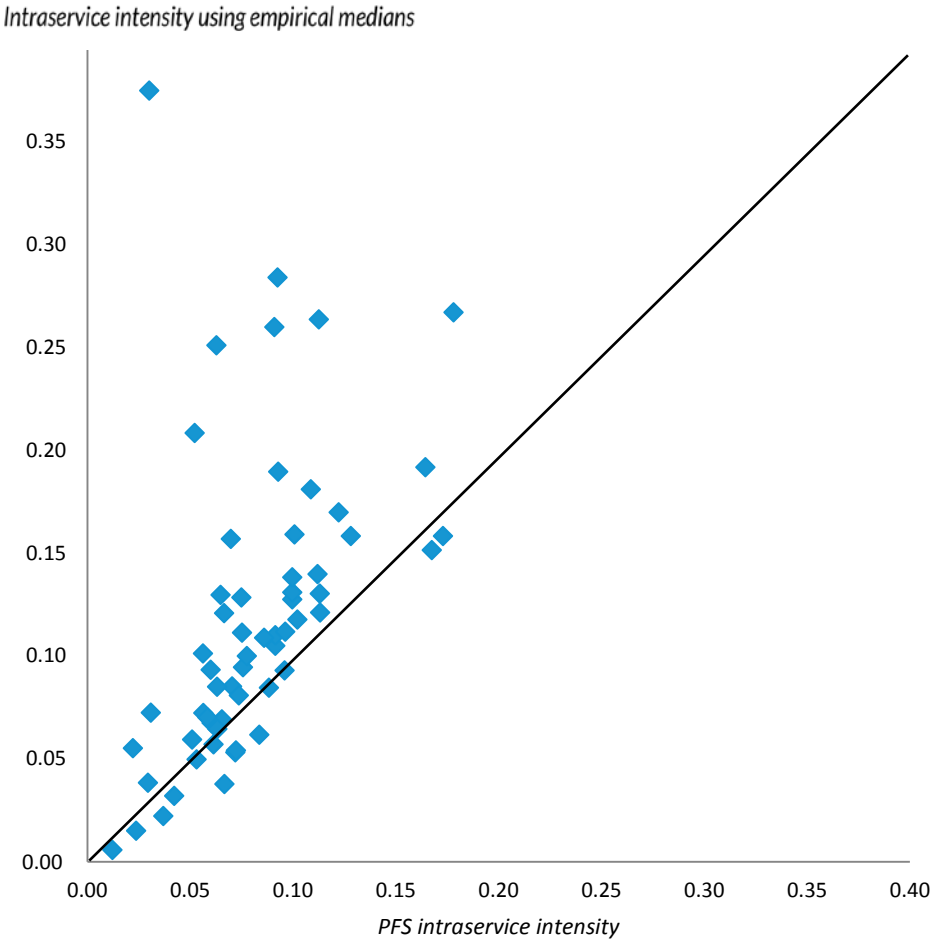
For this analysis, we assumed physician intraservice work was given and examined the implications of our alternative empirical intraservice time measure for intensity. If the implied intraservice intensity values seem incongruous with the current understanding of intensity, then either (1) intraservice work values are relatively accurate and the empirical intraservice time values are wrong; (2) the empirical intraservice time values are accurate and intraservice work values are wrong; or (3) some combination of (1) and (2).[10] In the analysis that follows, we compare intraservice intensities derived from study

intraservice time data to those derived from current PFS time data, and we compare *relative* intraservice intensities under the two measures.

Figure 1 shows the relationship between current implied intraservice intensity and intensity based on our median empirical intraservice time estimates, assuming the current PFS work is accurate for the sake of comparison.[11] The 45-degree line shows where the two values would be equal. Most implied intraservice intensities based on study intraservice times are above the line, indicating that this intraservice intensity (on the vertical axis) is higher than intraservice intensity under current PFS time estimates (on the horizontal axis). This is because our median intraservice time estimates are generally lower than the corresponding PFS values. For a given work RVU, lower service times correspond to higher intensities.

FIGURE 1

PFS Intraservice Intensity versus Intensity Using Empirical Medians at the HCPCS Level



Source: UI/SSS analysis of primary data and PFS 2016 Final Rule (80 FR 70885) public use files.
Note: One code with intensity near 1.5 (based on the empirical time estimate) was omitted from the chart to preserve the scale.

Across the study's 60 HCPCS codes, the median intraservice intensity using the study times was 0.11 (table 7), more than 50 percent higher than the 0.07 median intraservice intensity under PFS times.[12] Either these generally higher intraservice intensities are indefensible, or they suggest that some current intensities have been understated. If the study-based intensity values seem wrong, then the empirical intraservice time values collected in this project imply potential problems with the current work RVUs.

TABLE 7

Median Implied Intraservice Intensity under PFS and Empirical Intraservice Times, by Type of Service

	Median implied intra-service intensity N	Median implied intra service intensity using PFS time	Median implied intra-service intensity using empirical time	Median rank of PFS intensity	Median rank of study-based intensity
All types of service	60	0.07	0.11	--	--
Type of service					
Physician office-based procedures (procedures substantially performed in an office setting)	6	0.10	0.12	12.5	26
Outpatient department/ambulatory surgical center procedures, with or without global period (procedures substantially performed in OPD/ASC setting)	19	0.09	0.13	19.0	20
Inpatient procedures with global period	23	0.07	0.08	32.0	41
Imaging and other test interpretations	12	0.06	0.14	47.5	19

Source: UI/SSS analysis of primary data and PFS 2016 Final Rule (80 FR 70885) public use files.

If all intraservice intensity values were consistently larger under our time estimates by a similar percentage, then they would not affect relative intraservice work values across services.[13] Across the study's 60 HCPCS codes, however, the variation in the PFS intraservice intensity measure explains less than 40 percent of the variation in the study time-based intensity measure.[14] This suggests that there are relative shifts in intraservice intensity between the two measures. This raises two questions:

- Does the relationship between PFS- and study-based intraservice intensities differ for specific groups of services?
- To what extent does the position of specific codes relative to one another differ under the two measures of intraservice intensity?

Although the overall median intraservice intensity increases from 0.07 to 0.11 for all service types, the change in median intraservice intensity varies across specific types (table 7). The difference for both office-based procedures and global surgical procedures is much more modest, with the median intraservice intensity for each service category increasing only 0.01 between the two measures. Notably, imaging and interpretation services have the lowest median intraservice intensity under PFS time values and the highest intensity under study-based time values. The change in median intraservice intensity for OPD/ASC services is similar to the overall increase. These larger median intraservice

intensity changes for OPD/ASC and imaging and interpretation services are consistent with the fact that these two categories had the largest drop in intraservice times based on the empirical study data relative to PFS values (table 6).

To examine how relative intraservice intensities change, we calculated the rank of each service under the two intraservice intensity estimates. A low rank denotes a service with a relatively high intraservice intensity, while a high rank denotes a service with a relatively low intensity. The change in the median rank of the two intraservice intensity measures across the four categories, as shown in table 7, mirrors the change in the median intensity. The median rank of imaging and interpretation services is much higher under the empirical intraservice time measure than under the current PFS values. The median intraservice intensity rank for this category moves from 47.5 under PFS values to a median rank of 19 under the study measure, reflecting the substantially lower intraservice physician time.

Two other categories—office-based procedures and inpatient procedures—increase in rank when we use study times instead of PFS times, because these service categories exhibited the smallest differences between PFS and study intraservice times. The median rank of office-based services increases from 12.5 to 26, while that of inpatient procedures increases from 32 to 41; for both of these service categories, relative intensity declines. We observed little change in the median rank for OPD/ASC procedures, since the difference between median PFS and study intraservice times for this category was roughly equal to the overall change for the study's 60 HCPCS codes.

This suggests that the empirical intraservice time data had the largest relative effect on imaging and other test interpretations, for which study times tend to be much lower than current PFS values. Taking PFS work values as given, these lower times would imply that the intraservice intensity of these services is much higher than currently assumed. Put differently, the current intraservice intensity values for imaging and other test interpretations are the least related to our study-based values among the four categories.[15] OPD/ASC service intensity increases under the study intraservice time estimates just enough to be relatively unaffected by the increase in intensity for imaging and other test interpretations, while the intensity values for the other two categories—office-based procedures and inpatient procedures—fall relatively, given the smaller difference between PFS and study times for these services.

To illustrate the HCPCS code-level effect of study time on intraservice intensity, table 8 shows two groups of services. The nine HCPCS codes in the top panel all have intraservice intensity of about 0.07 under current PFS time values. However, under study intraservice time values, their intraservice intensities range from 0.04 for partial removal of colon (HCPCS code 44140) to 0.16 for interpretation

of brain stem MRI without dye (HCPCS code 70551).[16] These HCPCS codes have similar intraservice intensities under current PFS times but intensities that differ by a factor of 4 under our study time intensities. The bottom panel shows five HCPCS codes that have an intraservice intensity of 0.13 under our empirical times but have intraservice intensities that range from 0.06 (HCPCS code 70450, CT of the head/brain without dye) to 0.11 (HCPCS code 33430, replacement of the mitral valve) under the PFS times.

While the specific intraservice intensity estimates may not be precise, these two groups of services reveal large shifts in relative intraservice intensity under the two sets of time estimates. The upper panel forces us to consider which explanation is more reasonable for these two services:

1. Partial removal of the colon and interpretation of an MRI have similar intraservice intensities, as implied by current PFS times, or
2. Interpretation of an MRI has an intensity four times higher than partial removal of a colon.

The bottom panel poses a similar question: Do the intraservice intensities of a CT of the head/brain without dye and replacement of the mitral valve differ by roughly a factor of 2, as under the PFS intraservice time values, or are they roughly equivalent, as suggested by the empirical intraservice time estimates? These types of differences raise important questions about relative intraservice intensity under current PFS values, which could in turn have implications for relative work values.

If the empirical intraservice time estimates collected in this study better reflect clinical practice than those currently used in the PFS, then either our current understanding of intraservice intensity is wrong or relative intraservice work values are wrong. Consider, for example, two other HCPCS codes from the top panel of table 8: treatment of thigh fracture (HCPCS code 27244) and fragmentation of kidney stone (HCPCS 50590). Under the PFS, the implied intraservice work value for the thigh fracture treatment is 4.908, which is about 8 percent higher than the implied value of 4.496 for kidney stone fragmentation. Their PFS service times differ even more substantially (75 minutes and 60 minutes, respectively), so their intraservice intensities are also different, but both round to 0.07. However, the empirical intraservice time estimates for these two HCPCS codes show a larger difference than the PFS time values: while the thigh fracture drops from 75 to 71 minutes, kidney stone fragmentation drops from 60 to 35 minutes. These two intraservice time estimates support two alternative interpretations:

1. The intraservice intensity estimates for these two HCPCS codes should *not* be roughly equal, as they currently are, but in fact differ by nearly 100 percent, as they would using our empirical time estimate; or

2. The current intraservice intensity values are about right for these HCPCS codes and the implied intraservice work values for the two should differ more than they do currently. Based on the product of current intraservice intensity and the new empirical intraservice time estimates, new intraservice work RVUs would be 4.65 for thigh fracture treatment and 2.62 for kidney stone fragmentation.

Either interpretation has implications for existing intraservice work RVUs. The first suggests that using current intraservice intensity in the process of establishing or refining work RVUs should be reconsidered. The potential inaccuracy of current intraservice intensity values also raises questions about pre- and postservice intensity values for these services; these values may need to be recalculated.

The second interpretation would support a revision to intraservice work RVUs. However, if intraservice work RVUs change—under the assumption that pre- and postservice work are correctly captured in the building block method—total work would decrease by the same *number* of RVUs as intraservice work. However, the relative effect on total work may be very different than on intraservice work, since intraservice work represents a different share of total work for each service. In the case of these two HCPCS codes, intraservice work is about one-quarter of total work for thigh fracture treatment while it is roughly half of total work for kidney stone fragmentation.

TABLE 8

Comparison of Implied Intraservice Physician Time Based on PFS and Empirical Time for Select Services

HCPCS code	Brief service descriptor	Implied PFS intraservice work RVUs	Intra-service time: PFS	Intra-service time: empirical	Implied intra-service intensity using PFS time	Implied intraservice intensity using empirical time
	Panel 1: PFS implied intraservice intensity = 0.07					
44140	Partial removal of colon	9.9985	150	265.0	*0.07*	0.04
44143	Partial removal of colon	10.7985	150	203.0	*0.07*	0.05
44145	Partial removal of colon	12.9985	180	240.5	*0.07*	0.05
27244	Treat thigh fracture	4.9080	75	71.0	*0.07*	0.07
47563	Laparoscopic cholecystectomy/graph	6.6285	90	82.0	*0.07*	0.08
47562	Laparoscopic cholecystectomy	5.6285	80	66.0	*0.07*	0.09
27134	Revise hip joint replacement	15.9580	240	132.0	*0.07*	0.12
50590	Fragmenting of kidney stone	4.4961	60	35.0	*0.07*	0.13
70551	MRI brain stem w/o dye	1.2560	18	8.0	*0.07*	0.16
	Panel 2: Study implied intraservice intensity = 0.13					
70450	CT head/brain w/o dye	0.6484	10	5.0	0.06	*0.13*
50590	Fragmenting of kidney stone	4.4961	60	35.0	0.07	*0.13*
45378	Diagnostic colonoscopy	2.4907	25	19.5	0.10	*0.13*
45380	Colonoscopy and biopsy	2.7907	28	21.3	0.10	*0.13*
33430	Replacement of mitral valve	26.2120	232	201.0	0.11	*0.13*

Source: UI/SSS analysis of primary data and PFS 2016 Final Rule (80 FR 70885) public use files.

An important caveat to this analysis is that it takes as accurate and appropriate the pre- and postservice time, the pre- and postservice fixed intensity assumed across HCPCS codes in the building block method calculation of intraservice work and intensity, and the work associated with the E&M services included in the global period. If, for example, pre- and postservice times are inaccurate, they may be masking or exacerbating errors in the implied intraservice work value, which in turn would lead to erroneous implied intraservice intensity estimates for both current PFS time values and study time estimates. However, the findings of this pilot study suggest that PFS intraservice time may be distorted in ways that distort work RVUs, with the bulk of the distortions within the study's 60 HCPCS codes indicating that current intraservice work RVUs are too high.

We also cannot generalize the category-specific effects beyond the studied HCPCS codes. The specific HCPCS codes studied are important, high-volume services within the Medicare program and

merit close scrutiny. However, we do not know if the findings related to the study's 60 HCPCS codes reflect broader systematic bias in current PFS RVUs across other service categories.

Clinical Expert Review Results

This section summarizes what we learned from our semistructured interviews about the representativeness of vignettes and the accuracy of service descriptions. Findings from the interviews are discussed below, with a focus on reporting discrepancies and inaccuracies. We also examine clinical reviewers' intraservice time estimates.

Representativeness of Vignettes

Respondents generally agreed that vignette representativeness was an important factor in their ability to estimate time and work associated with specific services.

Across the families of codes we studied, clinical reviewers found the PFS vignettes to be fairly representative of typical patients for whom the HCPCS codes applied. Out of 25 vignettes, respondents only raised concerns about the representativeness of five vignettes, which are discussed below.

The vignette for prostatectomy (HCPCS code 55866) presents a 48-year-old patient. Respondents questioned the age of this patient, reporting that the typical patient for this code is likely to be older and covered by Medicare. A younger patient, likely to have fewer comorbidities, would have shorter time and work estimates—particularly for preservice elements—than a more typical, older patient. Respondents also thought that for transurethral resection of the prostate, or TURP (HCPCS code 52601), the 76-gram prostate in the vignette was atypically large and likely to produce inflated time estimates, since respondents agreed that this procedure's time and work are proportional to the size of the prostate.

Two vignettes presented complex patients that were not extremely unusual but required more work than the typical case. The PFS vignettes for a chest X-ray (HCPCS code 71020) and brain MRI with or without contrast (HCPCS code 70553) presented patients with known cases of cancers that clinical reviewers felt would require additional time to review and interpret, mostly because of the need to search for and document the presence or absence of metastases and other cancer complications.

The vignette for cystourethroscopy with fulguration (including cryosurgery or laser surgery) (HCPCS code 52224) created confusion among the respondents. The PFS vignette presents a patient with carcinoma of the bladder who had undergone routine prior surveillance cystourethroscopy and had a number of lesions fulgurated and biopsy samples collected. Respondents split over whether the vignette described a procedure that would typically be an initial procedure, rather than a follow-up, and over whether the intraservice elements described a procedure usually performed in an outpatient department or ambulatory surgical center as opposed to a physician's office.

Accuracy of Service Descriptions

Preservice. We selected two HCPCS codes from each of the six specialties (except ophthalmology) to discuss pre- and postservice descriptions in addition to intraservice elements. Overall, physicians agreed that the activities listed in the PFS preservice descriptions were performed by physicians or by nonphysician providers for many services, particularly procedural ones. However, they reported that many of the preservice elements are often performed by clinical staff without physician or nonphysician practitioner involvement. Further, for some HCPCS codes, the preservice work was typically performed in a prior or concurrent office visit with separate billing.

Of the 14 preservice descriptions discussed, respondents took issue with the details of eight descriptions. This information led us to question the accuracy of those descriptions. Two descriptions included service elements that were exaggerated. The remaining six descriptions included service elements that did take place but were typically performed at a prior or concurrent office visit with separate billing. Below, we discuss in more detail the issues raised by clinical reviewers regarding these eight descriptions.

Our respondents indicated that unless service elements involve clinical review of indications for the procedure or reviewing of imaging and lab studies, nonphysician providers assist physicians to a great degree, either in the prior office visit or on the day of the procedure. In other situations, clinical staff (including medical assistants and trained technicians) perform the identified physician tasks. Preservice elements performed by clinical staff ranged from operating room assistance on the day of the procedure to administrative activities during a prior office visit (e.g., obtaining informed consent, updating the patient's medical records, educating the patient and family about the procedure and the recovery phase, monitoring scheduling, and ensuring that the relevant equipment and staff would be available for the procedure).

For common procedures such as total hip arthroplasties (HCPCS code 27130), knee arthroscopies (HCPCS code 29881), prostate biopsies (HCPCS code 55700), prostatectomies (HCPCS code 55866), endoscopies (HCPCS code 43239), and colonoscopies (HCPCS code 45385), many of the preservice elements in the service descriptions are performed at an office visit a few days to a few weeks prior to the procedure. On the day of the procedure, most physicians meet the patient only to briefly review the latest imaging or lab results and, most importantly, to mark the correct site of operation and to answer any questions the patient may have.

Respondents reported that review of patient records and clinical indications—part of the service description for computerized optical coherence tomography (OCT) imaging studies (HCPCS codes 92133 and 92134)—occurred during a concurrent office visit and was not a preservice element specifically associated with the test. Thus, preservice time may be misestimated because of overlapping work for two simultaneously performed services billed with two service codes.

The service descriptions for thorax CT (HCPCS code 71250) and spinal cord MRI (HCPCS code 72148) mention that the physician should determine and communicate appropriate protocols to the imaging technician prior to the study. But the clinical reviewers reported that it is now common to use established electronic, predesigned protocols for the specific imaging studies to be performed. Though physicians periodically develop and refine these protocols, the technician is able to conduct the study without specific physician guidance in the vast majority of cases.

Moreover, a radiologist typically reviews prior studies while interpreting the new study as part of the *intraservice* work, rather than separately as preservice work (per the service description). Some respondents noted that reviewing the appropriateness of ordered imaging studies was missing from the service description. This preservice element sometimes, but atypically, consumes considerable time, as technicians communicate with the referring physician to arrive at the correct study to be performed.

Intraservice. Respondents generally agreed with the intraservice descriptions. The few areas of disagreement mostly stemmed from changes in technology or inaccuracies surrounding activities no longer performed. Of the 25 intraservice descriptions discussed, respondents raised significant concerns about the details of eight descriptions. This information led us to question the accuracy of those descriptions, described below.

The nature of intraservice work for interpretations of noninvasive cardiac testing, including electrocardiograms (HCPCS code 93010), echocardiograms (HCPCS code 93306), and cardiac stress tests (HCPCS code 93015), has been altered by both automation and personnel substitution. Physicians now receive the electrocardiogram tracings on their computer, and they consider the automated,

computer-embedded set of measurements to be reliable. The service description expects that the "primary intervals are measured and compared with computer-generated intervals, if available," but these computer-generated measurements and intervals are now ubiquitous. Findings from the interviews indicate that the interpreting physician rarely measures intervals using calipers to confirm the computer reading. Cardiologists routinely review the automated interpretations and change them if inaccurate; the computer-generated report is part of the official record. Clinical reviewers commented that the time for interpretation and the accuracy of automated interpretations varied depending on the source of the tracing: preoperative, emergency room, and ambulatory electrocardiograms (EKGs) tend to be straightforward, with mostly accurate automated interpretations, whereas cardiac care and intensive care unit EKGs tend to require more interpreting time and, sometimes, corrections to the automated report.

The cardiologists we interviewed agreed that for echocardiograms, technicians, not physicians, obtain a sequence of real-time tomographic images of cardiac structure and dynamics from multiple views, and technicians record the clips digitally, as listed in the service description of physician intraservice work. Technicians perform the entire study; cardiologists rarely perform it, as the service descriptions assume.

The respondents indicated that the nature of intraservice work for cardiac stress tests also varied with the availability of nurse practitioners (NPs) and physician assistants (PAs). The service descriptions list activities such as discussing recent history with the patient, examining the patient, assessing new symptoms since the test was ordered, and assessing adequacy of data collection; all of these tasks are now commonly performed by NPs and PAs. Clinical reviewers informed us that while these nonphysician providers perform the test, the physicians were always available in close proximity. In most cases, physicians reviewed the findings reported by the NP or PA and, when indicated, added to or corrected the interpretations. Only one interviewed cardiologist reported performing the entire test when visiting rural centers where trained staff were not usually available. It should be noted that the work performed by NPs, PAs, and other practitioners authorized to bill Medicare is considered physician work in the PFS. Even so, the widespread use of nonphysician providers may affect judgments about the work or intensity of this service.

Technological developments such as picture archiving and communication systems (PACS) have also changed the nature of the intraservice elements in radiology, although respondents disagreed over the extent of the technology's impact on time and work. Some respondents reported that interpretations had become easier and more efficient through PACS, while others observed that the

availability of easily obtainable images increased the number of images they had to review, with concomitant expectations of more findings.

The service description also fails to capture the impact of new technology on transurethral resection of the prostate (HCPCS code 52601). A relatively new bipolar technology allows for less blood loss and fewer postoperative complications, but it can make the procedure comparatively slower to perform.

In the discussion on preservice elements, we briefly touched upon the example of OCT imaging (HCPCS codes 92133 and 92134) preservice work being performed concurrently with an office visit. That overlap holds true for OCT intraservice elements as well. Clinical reviewers told us that not only do they interpret the OCT images during the office visit, they also discuss the results with the patient during the concurrent office visit; this discussion may be considered intraservice for OCT imaging. One respondent noted that other tests such as gonioscopy (a baseline visual field test) and Heidelberg retinal tomography are also usually done alongside OCT imaging and are interpreted simultaneously during the same concurrent office visit, introducing the possibility of double-counting time.[17]

Postservice. In most cases, electronic communication and assistance from clinical staff has helped make immediate postservice work more efficient for physicians. Of the 14 postservice descriptions discussed, clinical reviewers raised significant concerns about six descriptions. This information led us to question the accuracy of those descriptions, discussed below.

One major issue was the active involvement of clinical and other support staff. For procedures in orthopedics, urology, and gastroenterology, such as knee arthroscopies (HCPCS code 29881), total hip arthroplasties (HCPCS code 27130), prostate biopsies (HCPCS code 55700), prostatectomies (HCPCS code 55866), endoscopies (HCPCS code 43239), and colonoscopies (HCPCS code 45385), staff, rather than physicians (per the service descriptions), transfer the patient from the operating table or procedure room to the recovery area. For procedures such as prostate biopsy, which are typically carried out in a physician's office, patients simply "get up, get dressed, and leave." This differs from the service description, which indicates that the physician assists in the transfer of the patient from the operating table to the postoperative stretcher and the recovery area.

Respondents said that informing the referring physician after the procedure—considered a postservice element for most services we studied—is commonly automated through the EHR, unless the operating or interpreting physician makes a significant and/or urgent finding, in which case the referring physician would be notified, often by phone. For example, when interpreting imaging and test results for a patient in the ER, physicians would immediately report back to the referring physician rather than

relying on the EHR to push out the results to the referring or other interested physicians and nonphysician providers.

Intraservice Time Estimates

Given the small number of respondents who provided intraservice time estimates as part of the clinical review process, this analysis of their estimates is suggestive but not conclusive. Overall, their estimates reinforce our empirical time estimates' implication that current PFS time values are generally high. For some services there was substantial variation in respondents' time estimates for intraservice work, whereas for other services, estimates were fairly consistent (table 9). For example, estimates for inpatient procedures were much more consistent than those for office procedures and test interpretations, which displayed substantial variation. Based on our conversations with clinical reviewers, we concluded that a range of factors contribute to their estimates of intraservice time, including variation in practice support and potential conflation of concurrent office visits with the intraservice work of the specific test interpretation or procedure.

TABLE 9

Clinical Expert Reviewers' (CER's) Intraservice Time Estimates

Specialty	HCPCS code	Brief service descriptor	2016 PFS intra-service time (min)	Empirical time: median intraservice time (min)	Empirical time: N	CER esti-mate: 1 min[a]	CER esti-mate: 2 min[a]	CER esti-mate: 3 min[a]	CER esti-mate: 4 min[a]	CER esti-mate: 5 min[a]
Noninvasive cardiology	93010	EKG	5	0.1[b]	411	1	1	1	3	5
Noninvasive cardiology	93306	Echocardiogram	20	5	20	10	5	8	13	15
Gastroenterology	43235	EGD w/brush wash	15	5	61	15	13	15	8	10
Gastroenterology	43239	EGD w/biopsy	15	6	227	15	15	20	8	10
Gastroenterology	45378	Colonoscopy w/brush wash	25	20	304	38	20	30	18	23
Gastroenterology	45380	Colonoscopy w/biopsy	28	21	333	38	27	33	18	23
Gastroenterology	45385	Colonoscopy w/polyp removal	30	22	120	38	30	30	18	23
Ophthalmology	66984	Cataract surgery	21	18	1565	--	18	28	--	8
Ophthalmology	67028	Injection eye drug	5	N/A	<8	13	18	11	12	--
Ophthalmology	92133	OCT (optic nerve)	10	N/A	<8	8	13	2	9	4
Ophthalmology	92134	OCT (posterior segment)	10	N/A	<8	8	13	10	12	4
Orthopedics	27130	Total hip arthroplasty	100	87	471	83	90	95	75	115
Orthopedics	27236	Thigh fracture treatment	90	81	32	75	75	90	75	100
Orthopedics	27447	Total knee arthroplasty	100	83	726	73	85	85	75	110
Orthopedics	29881	Knee arthroscopy	40	N/A	<8	25	48	30	38	53
Radiology	70553	MRI brain w/ or w/o contrast	25	N/A	<8	13	17	23	20	14
Radiology	71020	X-ray chest	3	3	8	4	3	3	3	3
Radiology	71250	CT thorax w/o contrast	15	16	11	10	14	15	8	12
Radiology	72148	MRI spinal canal w/o contrast	20	N/A	<8	10	14	23	18	24
Urology	52000	Cystourethrosco-py	15	14	54	15	2	10	10	20
Urology	52224	Cystourethrosco-py w/ fulguration	30	19	45	33	20	40	20	30
Urology	52601	TURP	75	56	142	95	75	90	60	60
Urology	55700	Prostate biopsy	15	13	30	21	14	18	11	20
Urology	55866	Prostatectomy	180	176	95	180	180	190	180	270

Source: UI/SSS analysis of clinical review responses and PFS 2016 Final Rule (80 FR 70885) public use files.
Notes: [a] Physician time estimates (min) that were reported as ranges are shown as the midpoint of that range. [b] The median empirical intraservice time was 6 seconds.

For inpatient procedures, clinical reviewer time estimates were similar to both PFS time estimates and study time estimates. For some of these services, however, the observed consistency of intraservice

times does not validate total service time values because postservice activities account for a substantial share of total work.

For office procedures and test interpretations, clinical reviewer time estimates generally fell between empirical study times and PFS times. For example, almost all of the expert panel time estimates for the cardiac test interpretations were substantially lower than PFS times and closer to the empirical times. In only one case (prostate biopsy, HCPCS code 55700) did a majority of clinical reviewers (three out of five) provide a time estimate higher than the current PFS value.

Clinical reviewers did not think the vignettes or service elements were accurate for some services and provided time estimates based on their own understanding of the service. As noted earlier, 20 percent of the vignettes used atypical examples and, as a result, their associated time and work estimates were thought to be atypical. Respondents felt that most but not all of these atypical vignettes implied that the cases described required more time and work effort than a typical case.

BOX 1

Limitations: Challenges of Empirical Time Data Collection

In this pilot effort, we learned that collecting empirical data on physician time is feasible but far more complex than originally anticipated. Data collection requires continuous buy-in among clinical and administrative leadership throughout the organization. In the health systems we studied, leadership granted consent for participation but were not adequately engaged in data collection. Without ongoing leadership engagement, the process of initiating data collection can drag on or simply never happen. Beyond the three systems that were willing to provide data, we had initial contact with many others that expressed strong interest but never followed through. In addition, we encountered a variety of low-level challenges in conducting on-site data collection, including working with union rules, seeking IRB approval, obtaining patient or physician consent for direct observation, and determining where we could work without impeding the clinical workflow. Even though we offered to pay practices for participating in this study, it was surprisingly difficult for them to assign the staff time and other resources required for this type of project to develop a budget and subcontract.

A number of issues surrounding direct observation arose in the field and warrant consideration in future efforts of this kind. First, the scheduled service may not match the actual service provided, so it is difficult to plan direct observation for a targeted list of HCPCS codes. This challenge was so significant that we would recommend organizing future direct observation efforts around logging the activities of a set of physicians within a specific time period and then identifying the services they provide, as opposed to starting with a set of targeted services. Second, health systems were sometimes reluctant to let project staff conduct the direct observation and wanted this done by their own clinical staff. This

required us to train practice personnel well enough that they could distinguish between specific aspects of each service and between different types of clinical personnel. Although training was feasible, our inability to observe the actual data collection made oversight and quality control difficult. Third, it was difficult to collect pre- and/or postservice time using direct observation of specific services, since for some tasks, the pre- and postservice physician tasks can happen at very different times or places than the intraservice tasks. Thus, the study could only focus on intraservice work. The issues related to observing pre- and postservice work could be overcome by using the same approach we suggested for general service tracking—namely, the preliminary logging of physician activities.

Working with data from EHRs was also difficult but for very different reasons. First, not only do EHR systems vary across practices, they may also be employed inconsistently within a given practice. We found that different departments within a practice may employ different features of a particular EHR system and, in some instances, different departments may employ entirely different systems. For example, inpatient operating room suites may use EHR features that are different from those used in ambulatory surgery centers. This sometimes made it difficult to collect similar data across a range of services. Second, EHR data are likely to be available for certain types of services and not for others (e.g., office-based procedures or tests), and direct observation is feasible for some types of services and not for others. This dissimilarity complicated the comparison of time estimates derived from these two alternative empirical approaches. Third, sites that do not rely on or use the data required for this study (for example, service times or HCPCS codes) are unlikely to provide helpful data. We found that practices that do not rely on HCPCS codes for fee-for-service reimbursement do not track HCPCS codes particularly well for all services.

Additional details related to the challenges of data collection are presented in appendix C.

Summary

After overcoming the challenges of identifying, recruiting, and working with three multispecialty health care systems, we were able to develop empirical physician time data for 60 HCPCS codes. Our experience suggests that the MedPAC contractor report assessing the feasibility of this task was overly optimistic; that report reviewed the early steps of this process in a largely hypothetical context (Braun and McCall 2011). As we moved toward obtaining formal commitments from health systems, engaging with providers within these organizations, and estimating the time required from many of the administrators and data support staff, it became clear that the earlier study had understated the hurdles involved in this process. In response to these challenges, we modified our initial project design along the way (Zuckerman et al. 2014).

Based on the data we collected, we concluded that PFS intraservice physician time was often high relative to the empirical time captured from our study. For 42 of the 60 HCPCS codes we studied, PFS intraservice physician time was more than 10 percent above this study's median intraservice physician time. The clinical expert review of a subset of the 60 HCPCS codes tended to confirm that the time values reflected in these new data were consistent with their experience. The largest intraservice time discrepancies we observed occurred in imaging and other test interpretations, outpatient department and ambulatory surgical center procedures, and office-based procedures. There were smaller differences among inpatient procedures, with many procedures seeing intraservice times below those in the PFS. There were, however, some inpatient procedures with median empirical intraservice time more than 20 percent above PFS times.

We examined the effect of these new time estimates on intraservice work intensity to consider how work RVUs might be distorted. Not surprisingly, taking intraservice work RVUs as given, our lower empirical time estimates suggest that current implied PFS intraservice intensity is often too low. More importantly, for a relative value scale, differences in intensity implied by these new time data are not uniform across services. However, if current intensity values seem more credible than those implied using the new time data, current work RVUs may be too high for many services. With respect to the new time information or the new intraservice intensities that it implies, this pilot study provides evidence that the relative work RVUs of many PFS services may be distorted, mostly on the high side.

Recommendations for Further Data Collection

Beyond the substantive implications of our new empirical time data, there were two important sets of lessons we learned from this pilot study. The first relates to the data collection process itself. The major data collection challenges we encountered included recruiting health systems and engaging their staff, collecting intraservice data for a prespecified set of services, observing pre- and postservice time, and uniformly accessing the range of EHR data systems in use. Additional data collection issues are detailed in appendix C. These challenges informed the following guidance for future efforts to develop empirical estimates of physician service time.

- A broader study to collect empirical time data for work RVU validation seems feasible as long the health systems make this a high priority and ensure that adequate resources are available. The critical first step in recruiting health systems will be getting a sufficient number of senior-level staff to buy into the process so that other staff will engage effectively with data collection. This may require the involvement of CMS senior leadership to assist with recruitment and to commit substantially more resources than we had available for the health systems in this project. Based on our experience, we would suggest working with a manageable number of health systems rather than a broadly representative sample of physician practices. Small samples can be suitable as long as they reflect the diversity of methods of physician compensation across geographic areas and academic health center affiliations.
- Instead of starting with targeted HCPCS codes, future direct observation efforts should log the activities of a set of physicians within a specific time period and then identify the HCPCS codes of the services provided. This would require identifying the tasks provided by nonphysician providers (including NPs and PAs) who can bill for services under the PFS. Our experience in this project indicates that distinguishing between the tasks performed by physicians, nonphysician providers, and clinical staff would probably add to observation costs and require sophisticated clinical reviewers.
- Linking empirical times to individual HCPCS codes is not straightforward for either direct observation or EHRs. Frequently, multiple HCPCS codes are recorded for a patient on a single day, making it difficult to allocate minutes to specific HCPCS codes. Outside of a procedure room, physicians also commonly multitask to save time. In some instances, the preservice work

for one HCPCS code may be provided concurrently with the intraservice work of another HCPCS code. Finally, an observer may not know the final HCPCS code(s) that will be billed for a specific patient-provider encounter.

- Subsequent efforts to collect empirical time data should include E&M services, especially for office-based procedures and tests that occur during visits.
- No single mode of data collection will work well for all services. EHRs are likely to be useful for services that are provided in a dedicated place (like an operating room), do not typically involve interruptions, and are otherwise hard to observe. For other services, direct observation may be both feasible and essential. Direct observation is necessary for services that are not tracked accurately with time stamps in EHR systems, may involve multiple types of providers and clinical staff, and may only be completed after several interruptions. Since multiple strategies are needed across different services, there must be some way of calibrating the accuracy of different approaches when both can be used for the same services.

The second set of lessons derives from the clinical expert review, which focused in part on the accuracy and usefulness of the typical patient vignettes and service descriptions used by the RUC to enumerate the components of intraservice physician work and to delineate pre- and postservice activities.

- The clinical expert review provides a great deal of qualitative information about specific procedures and should be viewed as an integral part of any future data collection efforts. In fact, based on the insights we gathered, we would recommend expanding this component of the project to cover more services and more specialties. The design of the data collection strategy could benefit from conducting the clinical expert review at the beginning of the project as well as during the final analysis and interpretation of resulting time estimates and their implications for work RVUs.
- The RUC vignettes used in the specialty society surveys should be systematically reviewed to ensure that they describe a typical patient for each study service; this will help prevent bias in assigning time and work values. However, these vignettes are not essential when collecting empirical time data because that process measures time spent with actual patients. With large enough samples, the central tendency of the data would identify what is typical, but this still would be benchmarked against what an up-to-date RUC vignette views as typical. If physician surveys will continue to be used in the process of establishing or updating RVUs, then it is important to make sure that the vignettes accurately represent the typical patient. One

challenge of developing these vignettes, according to our clinical experts, is that there is natural variation in clinical presentation across patients.

- The detailed descriptions of physician activities included in the pre-, intra-, and postservice periods used by the RUC should be reviewed periodically and updated to conform to current clinical practice. About one-third of the 25 intraservice descriptions we discussed with our respondents were sufficiently problematic that we considered them inaccurate. Reviewers of the descriptions also noted that it was difficult to cleanly separate time and work for specific HCPCS codes when services are provided during concurrent patient visits. These conclusions are based on a small sample of HCPCS codes, but they indicate problems with the service descriptions that could affect PFS intraservice times and work RVUs. The distinctions between pre-, intra-, and postservice time must be clear to allow for proper mapping of EHR time and recording of observation data.

Appendix A. Services Selected for Study

Table A.1 displays the 117 services that met the original selection criteria we used when designing the data collection effort, as described in the interim report (Zuckerman et al. 2014).

TABLE A.1

117 Services Meeting the Original Selection Criteria, with Code Descriptors

HCPCS code	Code descriptor
11042	Debridement, subcutaneous tissue (includes epidermis and dermis, if performed); first 20 sq cm or less
11056	Paring or cutting of benign hyperkeratotic lesion (e.g., corn or callus); two to four lesions
11057	Paring or cutting of benign hyperkeratotic lesion (e.g., corn or callus); more than four lesions
11100	Biopsy of skin, subcutaneous tissue and/or mucous membrane (including simple closure), unless otherwise listed (separate procedure); single lesion
11101	Biopsy of skin, subcutaneous tissue and/or mucous membrane (including simple closure), unless otherwise listed (separate procedure); each separate/additional lesion (list separately in addition to code for primary procedure)
17000	Destruction (e.g., laser surgery, electrosurgery, cryosurgery, chemosurgery, surgical curettement), premalignant lesion (e.g., actinic keratoses); first lesion
17003	Destruction (e.g., laser surgery, electrosurgery, cryosurgery, chemosurgery, surgical curettement), premalignant lesion (e.g., actinic keratoses); 2 through 14 lesions, each (list separately in addition to code for first lesion)
17004	Destruction (e.g., laser surgery, electrosurgery, cryosurgery, chemosurgery, surgical curettement), premalignant lesions (e.g., actinic keratoses), 15 or more lesions
17110	Destruction (e.g., laser surgery, electrosurgery, cryosurgery, chemosurgery, surgical curettement) of benign lesions other than skin tags or cutaneous vascular proliferative lesions; up to 14 lesions
17262	Destruction (e.g., laser surgery, electrosurgery, cryosurgery, chemosurgery, surgical curettement), trunk, arms, or legs; lesion diameter 0.5 cm or less
17281	Destruction (e.g., laser surgery, electrosurgery, cryosurgery, chemosurgery, surgical curettement), face, ears, eyelids, nose, lips, mucous membrane; lesion diameter 0.6 cm to 1.0 cm
17282	Destruction (e.g., laser surgery, electrosurgery, cryosurgery, chemosurgery, surgical curettement), face, ears, eyelids, nose, lips, mucous membrane; lesion diameter 1.1 cm to 2.0 cm
20550	Injection(s); tendon sheath, or ligament, aponeurosis (e.g., plantar fascia)
20605	Arthrocentesis, aspiration and/or injection; intermediate joint or bursa (e.g., temporomandibular, acromioclavicular, wrist, elbow or ankle, olecranon bursa)
20610	Arthrocentesis, aspiration and/or injection; major joint or bursa (e.g., shoulder, hip, knee joint, subacromial bursa)
22551	Arthrodesis, anterior interbody, including disc space preparation, discectomy, osteophytectomy and decompression of spinal cord and/or nerve roots; cervical below C2
22612	Arthrodesis, posterior or posterolateral technique, single level; lumbar (with lateral transverse technique, when performed)
22614	Arthrodesis, posterior or posterolateral technique, single level; each additional vertebral segment (list separately in addition to code for primary procedure)
22633	Arthrodesis, combined posterior or posterolateral technique with posterior interbody technique including laminectomy and/or discectomy sufficient to prepare interspace (other than for decompression), single interspace and segment; lumbar

HCPCS code	Code descriptor
22840	Posterior nonsegmental instrumentation (e.g., Harrington rod technique, pedicle fixation across 1 interspace, atlantoaxial transarticular screw fixation) (list separately in addition to code for primary procedure)
22842	Posterior segmental instrumentation (e.g., pedicle fixation, dual rods with multiple hooks and sublaminal wires); 3 to 6 vertebral segments (list separately in addition to code for primary procedure)
22845	Anterior instrumentation; 2 to 3 vertebral segments (list separately in addition to code for primary procedure)
22851	Application of intervertebral biomechanical device(s) (e.g., synthetic cage(s), threaded bone dowel(s), methyl methacrylate) to vertebral defect or interspace (list separately in addition to code for primary procedure)
23412	Repair of ruptured musculotendinous cuff (e.g., rotator cuff) open; chronic
23472	Arthroplasty, glenohumeral joint; total shoulder (glenoid and proximal humeral replacement [e.g., total shoulder])
27130	Arthroplasty, acetabular and proximal femoral prosthetic replacement, with or without autograft or allograft
27134	Revision of total hip arthroplasty; both components, with or without autograft or allograft
27236	Open treatment of femoral fracture, proximal end, neck, internal fixation or prosthetic replacement
27244	Treatment of intertrochanteric, pertrochanteric, or subtrochanteric femoral fracture; with plate/screw type implant, with or without cerclage
27245	Treatment of intertrochanteric, pertrochanteric, or subtrochanteric femoral fracture; with intramedullary implant, with or without interlocking screws and/or cerclage
27447	Arthroplasty, knee, condyle and plateau; medial and lateral compartments with or without patella resurfacing (total knee arthroplasty)
29827	Arthroscopy, shoulder, surgical; with rotator cuff repair
33208	Insertion of new or replacement of permanent pacemaker with transvenous electrode(s); atrial and ventricular
33249	Insertion or replacement of permanent pacemaker with transvenous lead(s); single or dual chamber
33405	Replacement, aortic valve, with cardiopulmonary bypass; with prosthetic valve other than homograft or stentless valve
33430	Replacement, mitral valve, with cardiopulmonary bypass
33518	Coronary artery bypass, using venous graft(s) and arterial graft(s); two venous grafts (list separately in addition to code for primary procedure)
33519	Coronary artery bypass, using venous graft(s) and arterial graft(s); three venous grafts (list separately in addition to code for primary procedure)
33533	Coronary artery bypass, using arterial graft(s); single arterial graft
33536	Coronary artery bypass, using arterial graft(s); four or more coronary arterial grafts
35301	Thromboendarterectomy, including patch graft if performed; carotid, vertebral, subclavian, by neck incision
43235	Upper gastrointestinal endoscopy including esophagus, stomach, and either the duodenum and/or jejunum as appropriate; diagnostic, with or without collection of specimen(s) by brushing or washing (separate procedure)
43239	Upper gastrointestinal endoscopy including esophagus, stomach, and either the duodenum and/or jejunum as appropriate; with biopsy, single or multiple
44120	Enterectomy, resection of small intestine; single resection and anastomosis
44140	Colectomy, partial; with anastomosis
44143	Colectomy, partial; with end colostomy and closure of distal segment (Hartmann type procedure)
44145	Colectomy, partial; with coloproctostomy (low pelvic anastomosis)

HCPCS code	Code descriptor
44160	Colectomy, partial, with removal of terminal ileum with ileocolostomy
44204	Laparoscopy, surgical; colectomy, partial, with anastomosis
44205	Laparoscopy, surgical; colectomy, partial, with removal of terminal ileum with ileocolostomy
44207	Laparoscopy, surgical; colectomy, partial, with anastomosis, with coloproctostomy (low pelvic anastomosis)
45378	Colonoscopy, flexible, proximal to splenic flexure; diagnostic, with or without collection of specimen(s) by brushing or washing, with or without colon decompression (separate procedure)
45380	Colonoscopy, flexible, proximal to splenic flexure; with biopsy, single or multiple
45384	Colonoscopy, flexible, proximal to splenic flexure; with removal of tumor(s), polyp(s), or other lesion(s) by hot biopsy forceps or bipolar cautery
45385	Colonoscopy, flexible, proximal to splenic flexure; with removal of tumor(s), polyp(s), or other lesion(s) by snare technique
G0105	Colorectal cancer screening; colonoscopy on individual at high risk
47562	Laparoscopy, surgical; cholecystectomy
47563	Laparoscopy, surgical; cholecystectomy with cholangiography
49505	Repair initial inguinal hernia, age 5 years or over; reducible
50590	Lithotripsy, extracorporeal shock wave
52000	Cystourethroscopy (separate procedure)
52224	Cystourethroscopy, with fulguration (including cryosurgery or laser surgery) or treatment of minor (less than 0.5 cm) lesion(s) with or without biopsy
52281	Cystourethroscopy, with calibration and/or dilation of urethral stricture or stenosis, with or without meatotomy, with or without injection procedure for cystography, male or female
52601	Transurethral electrosurgical resection prostate, including control of postoperative bleeding, complete (vasectomy, meatotomy, cystourethroscopy, urethral calibration and/or dilation, and internal urtherotomy are included)
55700	Biopsy, prostate; needle or punch, single or multiple, any approach
55866	Laparoscopy, surgical prostatectomy, retropubic radical, including nerve sparing, includes robotic assistance, when performed
63047	Laminectomy, facetectomy and foraminotomy (unilateral or bilateral with decompression of spinal cord, cauda equina and/or nerve root[s] [e.g., spinal or lateral recess stenosis]), single vertebral segment; lumbar
64483	Injection, anesthetic agent and/or steroid, transforaminal epidural with imaging guidance (fluoroscopy or CT); lumbar or sacral, single level
66821	Discussion of secondary membranous cataract (opacified posterior lens capsule and/or anterior hyaloid); laser surgery (e.g., YAG laser) (one or more stages)
66982	Extracapsular cataract removal with insertion of intraocular lens prosthesis (one stage procedure), manual or mechanical technique (complex, requiring devices or techniques not generally used in routine cataract surgery [e.g., iris expansion device, suture support for intraocular lens, or primary posterior capsulorhexis] or performed on patients in the amblyogenic developmental stage)
66984	Extracapsular cataract removal with insertion of intraocular lens prosthesis (one stage procedure), manual or mechanical technique (e.g., irrigation or aspiration or phacoemulsification)
67028	Intravitreal injection of a pharmacologic agent (separate procedure)
67210	Destruction of localized lesion of retina (e.g., macular edema, tumors), one or more sessions; photocoagulation
67228	Treatment of extensive or progressive retinopathy, one or more sessions; (e.g., diabetic retinopathy), photocoagulation
70450	Computed tomography, head or brain; without contrast material

HCPCS code	Code descriptor
70486	Computed tomography, maxillofacial area; without contrast material
70551	Magnetic resonance (e.g., proton) imaging, brain (including brain stem); without contrast material
70553	Magnetic resonance (e.g., proton) imaging, brain (including brain stem); without contrast material, followed by contrast material(s) and further sequences
71010	Radiologic examination, chest; single view, frontal
71020	Radiologic examination, chest; two views, frontal and lateral
71250	Computed tomography, thorax; without contrast material
71260	Computed tomography, thorax; with contrast material(s)
71275	Computed tomographic angiography, chest (noncoronary), without contrast material(s), followed by contrast material(s) and further sections, including image postprocessing
72125	Computed tomography, cervical spine; without contrast material
72141	Magnetic resonance (e.g., proton) imaging, spinal canal and contents, cervical; without contrast material
72148	Magnetic resonance (e.g., proton) imaging, spinal canal and contents, lumbar; without contrast material
72158	Magnetic resonance (e.g., proton) imaging, spinal canal and contents, without contrast material, followed by contrast material(s) and further sequences; lumbar
74176	Computed tomography, abdomen and pelvis; without contrast material
74177	Computed tomography, abdomen and pelvis; with contrast material
74178	Computed tomography, abdomen and pelvis; without contrast material in one or both body regions, followed by contrast material(s) and further sections in one or both body regions
77080	Dual-energy X-ray absorptiometry (DXA), bone density study, one or more sites; axial skeleton (e.g., hips, pelvis, spine)
78452	Myocardial perfusion imaging, tomographic (SPECT) (including attenuation correction, qualitative or quantitative wall motion, ejection fraction by first pass or gated technique, additional quantification, when performed); multiple studies, at rest and/or stress (exercise or pharmacologic) and/or redistribution and/or rest reinjection
88305	Level IV - surgical pathology, gross and microscopic examination
88307	Level V - surgical pathology, gross and microscopic examination
88309	Level VI - surgical pathology, gross and microscopic examination; bone resection; breast, mastectomy - with regional lymph nodes; colon, segmental resection for tumor
88312	Special stain including interpretation and report; group I for microorganisms (e.g., acid fast, methenamine silver)
88331	Pathology consultation during surgery; first tissue block, with frozen section(s), single specimen
92133	Scanning computerized ophthalmic diagnostic imaging, posterior segment, with interpretation and report, unilateral or bilateral; optic nerve
92134	Scanning computerized ophthalmic diagnostic imaging, posterior segment, with interpretation and report, unilateral or bilateral; retina
92557	Comprehensive audiometry threshold evaluation and speech recognition (92553 and 92556 combined)
92920	Percutaneous transluminal coronary angioplasty; single major coronary artery or branch
92928	Percutaneous transcatheter placement of intracoronary stent(s), with coronary angioplasty when performed; single major coronary artery or branch
92941	Percutaneous transluminal revascularization of acute total/subtotal occlusion during acute myocardical infarction, coronary artery or coronary artery bypass graft, any combination of intracoronary stent, atherectomy and angioplasty, including aspiration thrombectomy when performed, single vessel
93000	Electrocardiogram, routine ECG with at least 12 leads; with interpretation and report

HCPCS code	Code descriptor
93010	Electrocardiogram, routine ECG with at least 12 leads; interpretation and report only
93015	Cardiovascular stress test using maximal or submaximal treadmill or bicycle exercise, continuous electrocardiographic monitoring, and/or pharmacological stress; with physician supervision only, with interpretation and report
93016	Cardiovascular stress test using maximal or submaximal treadmill or bicycle exercise, continuous electrocardiographic monitoring, and/or pharmacological stress; with physician supervision only, without interpretation and report
93018	Cardiovascular stress test using maximal or submaximal treadmill or bicycle exercise, continuous electrocardiographic monitoring, and/or pharmacologic stress; interpretation and report only
93306	Echocardiography, transthoracic, real-time with image documentation (2D), includes M-mode recording, when performed, complete, with spectral Doppler echocardiography, and with color flow Doppler echocardiography
93458	Catheter placement in coronary artery(s) for coronary angiography, including intraprocedural injection(s) for coronary angiography, imaging supervision and interpretation; with left heart catheterization including intraprocedural injection(s) for left ventriculography, when performed
93459	Catheter placement in coronary artery(s) for coronary angiography, including intraprocedural injection(s) for coronary angiography, imaging supervision and interpretation; with left heart catheterization including intraprocedural injection(s) for left ventriculography, when performed, catheter placement(s) in bypass graft(s) (internal mammary, free arterial, venous grafts) with bypass graft angiography
93460	Catheter placement in coronary artery(s) for coronary angiography, including intraprocedural injection(s) for coronary angiography, imaging supervision and interpretation; with right and left heart catheterization including intraprocedural injection(s) for left ventriculography, when performed
93880	Duplex scan of extracranial arteries; complete bilateral study
96372	Therapeutic, prophylactic, or diagnostic injection (specify substance or drug); subcutaneous or intramuscular
G0202	Screening mammography, producing direct digital image, bilateral, all views
G0204	Diagnostic mammography, producing direct digital image, bilateral, all views
G0206	Diagnostic mammography, producing direct digital image, unilateral, all views

Note: The code descriptors are based on those usually used by CMS. However, for clarity, words are spelled out completely in the version shown here.

Appendix B. Clinical Expert Review

This appendix provides more in-depth summaries of the interviews we conducted with physicians. As described in our report, we conducted interviews with five physicians in each of the six specialties. These physicians reviewed the 25 HCPCS codes selected as the focus of this clinical review. For each discussion, we provided respondents with each HCPCS code's vignette, describing a "typical" patient presenting for the service, along with the intraservice descriptions. We additionally selected two HCPCS codes in each specialty (all four in ophthalmology) for further discussion with respondents regarding the pre- and postservice descriptions. We used the RUC vignettes and service descriptions to provide the context for intraservice time estimates so that the clinical reviewers' time estimates could be compared to the RUC time estimates. This appendix is organized by specialty.

Noninvasive Cardiac Testing

Summary: We reviewed three noninvasive cardiac testing services with clinical reviewers: electrocardiogram (HCPCS code 93010), echocardiogram (HCPCS code 93306), and cardiac stress test (HCPCS code 93015). We discussed the accuracy of the preservice and postservice descriptions for electrocardiograms and echocardiograms and the accuracy of the intraservice descriptions for all three codes. Respondents indicated that the physician tasks involved in performing and interpreting these tests had changed because of improved technology with automated measurements and interpretations, and because of the enhanced roles played by clinical staff and nonphysician providers such as NPs and PAs.

Vignettes: Clinical reviewers found the vignettes for these three procedures to be representative, but they generally agreed that the tasks involved in these procedures rarely vary by patient characteristics.

Preservice descriptions: Clinical reviewers found the preservice descriptions to be accurate. For two services, electrocardiogram and cardiac stress test, the service descriptions list no activities for the preservice period. For echocardiograms, the clinical reviewers reported that technicians performing the test reviewed existing information and relevant clinical records to verify the indications for the procedure. The service descriptions include this as a physician task.

Intraservice descriptions: For the most part, respondents agreed with the descriptions of the intraservice tasks, with some important caveats. Physician tasks have changed with increased

automation, especially in two ways: it now takes less time to make interpretations, and the accuracy of the interpretations has improved.

Clinical reviewers agreed that automation has considerably changed the activities involved in interpreting EKG results. Four of five clinical reviewers reported receiving EKGs electronically and interpreting them online or, in one case, receiving a printed copy with computer-generated measurements on which the physician wrote his interpretation, which staff then entered into the EHR. Generally, respondents reported a high degree of confidence in computer-generated measurements and only verified those measurements in unusual circumstances. One respondent sometimes uses calipers to confirm computer-generated measurements.

Clinical reviewers indicated that they review the automated interpretations and make changes if required because the computer-generated report eventually becomes part of the patient's record. Reviewers commented that the time for interpretation and the accuracy of automated interpretations vary depending on the source of the tracing. Tracings generated from preoperative, emergency room, and ambulatory EKGs are commonly straightforward and mostly accurate automated interpretations, whereas tracings from cardiac care and intensive care units require more time and, sometimes, corrections to the automated report.

According to our clinical reviewers, technicians performed echocardiograms and recorded images for the physicians to review—contrary to the service description, which assumes the physician performs the test. Clinical reviewers informed us that only in rare circumstances do they actually record the echocardiogram images. They also reported variation in whether the technicians or the physicians do the measurements while reviewing the digitally recorded clips, but they agreed that intraservice tasks for physicians begin with interpretation of the digital clips.

Clinical reviewers also noted that they interpret and report echocardiogram findings simultaneously—contrary to the service description, which lists the task of preparing the report as a postservice activity. Including report preparation as a postservice activity is also inconsistent with service descriptions for EKGs and cardiac stress tests, which include the task of dictating (or preparing) the report as a part of the intraservice description.

Clinical reviewers observed that the availability of other staff to perform stress tests has largely reduced physician activity to test interpretation only. Clinical reviewers reported that nonphysician providers such as NPs and PAs perform the stress test and record the measurements, while physicians are available in close proximity in case their active supervision or clinical intervention is required. From Medicare's perspective, the time required to do these tasks by nonphysician providers counts as

physician time. Although two clinical reviewers interpret and report all findings, three clinical reviewers said they only review the report prepared by the NPs or PAs. In the event of a significant finding, the physician immediately follows up with the patient, but that is not typical. One clinical reviewer performs the entire stress test at rural centers that do not have staff available to assist.

Postservice descriptions: For EKGs, clinical reviewers reported that reviewing and signing the report is typically done electronically through the EHR with the click of a button. For echocardiograms, as indicated above, clinical reviewers interpret and report findings simultaneously, not as postservice tasks as detailed in the service description.

Intraservice time estimates: The intraservice time estimates from our respondents are presented in table 9 (in the body of the report). Most respondents estimated it takes about a minute or less to read and report an EKG, compared to the five minutes of intraservice time listed in the PFS. Only one respondent, who reads all EKGs on a printed copy rather than directly online, estimated a time of two to five minutes to interpret an EKG. Clinical reviewers indicated that the time taken to interpret and report findings for echocardiograms is five to ten minutes for typical cases, compared to the 20 minutes of intraservice time assumed in the PFS.

For cardiac stress tests, three clinical reviewers said that it takes two to five minutes to review the interpretation—compared with the 20 minutes listed in the PFS—and one clinical reviewer who personally interprets the results without assistance typically takes 10 minutes, with duration being a function of how long the patient is able to perform on the treadmill (for nonpharmacological stress tests). However, because the NP and PA time for conducting the test is considered to be physician time, it would be necessary to add together the separate NP and PA time with the physician interpretation time to arrive at the correct estimate of total physician time. We were unable to obtain estimates of NP and PA time using our interview protocol with clinical reviewers, so we are also unable to provide a time estimate for this service. The clinical reviewer who provides the entire stress test in rural centers estimated taking 15 to 20 minutes per patient.

Finally, one clinical reviewer suggested that there has been a relative increase in the proportion of pharmacological cardiac stress tests compared to classic treadmill tests, largely because the pharmacological test is being performed in older patients unable to exercise on the treadmill. Pharmacological stress tests are typically shorter in duration than treadmill tests.

Gastroenterology

Summary: We evaluated five upper and lower gastrointestinal endoscopy service codes: diagnostic endoscopy with specimen collection, brushing or washing (HCPCS code 43235); endoscopy with biopsy (HCPCS code 43239); diagnostic colonoscopy with specimen collection, brushing or washing (HCPCS code 45378); colonoscopy with biopsy (HCPCS code 45380); and colonoscopy with polyp removal (HCPCS code 45385). We discussed pre- and postservice description accuracy for endoscopy with biopsy and colonoscopy with polyp removal, and intraservice description accuracy for all codes. Clinical reviewers suggested that the vignette for diagnostic colonoscopy is inaccurate because it presents a 64-year-old patient who is referred for colorectal cancer screening, with no mention of findings that would alter the code to one of the diagnostic colonoscopy codes.

Clinical reviewers indicated that the intraservice descriptions are generally accurate, with one important exception. They pointed out an inaccuracy in the description of the technique used in polyp removal during a colonoscopy. The assumption in the service description that physicians routinely use retrieval devices and withdraw the endoscopes each time they remove a polyp is incorrect and can lead to inflated time estimates. Reviewers also observed that many of the preservice tasks in the service descriptions are generally performed at a prior office visit, not as part of the preservice period for the procedure.

Vignettes: Clinical reviewers found the vignettes to represent typical cases but made two other observations: For diagnostic endoscopy with single or multiple biopsies, physicians no longer do the *H. pylori* rapid urease test. For both endoscopies and colonoscopies, physicians no longer collect specimens by brushing or washing, per the service descriptions.

The RUC vignette for diagnostic colonoscopy presented a patient who is referred for colorectal cancer screening. Respondents said they would usually use a screening test G code while billing for a screening colonoscopy, but if they found abnormalities, removed a polyp, or collected a specimen for biopsy, they would bill for the appropriate colonoscopy code rather than the screening service.

Preservice descriptions: All clinical reviewers noted that most of the preservice tasks outlined in the service description, such as reviewing the patient's clinical history, physical exam, imaging studies, and other lab results, take place at a pre-assessment during a prior office visit; the physician only briefly reviews the patient's history, labs, and chart on the day of the procedure. Informed consent, listed as a preservice activity, is typically obtained by nurses or other office staff during a prior office visit and is rarely obtained on the day of the procedure.

The service description for endoscopy with biopsy states that the physician verifies that all endoscopic equipment is available and that operational and appropriate computer entries are made. On the contrary, most clinical reviewers informed us that their clinical staff performs these activities.

Intraservice descriptions: Overall, clinical reviewers agreed that intraservice descriptions for these services are accurate but made a few observations. For diagnostic endoscopy with specimen collection by brushing or washing and for diagnostic colonoscopy with specimen collection by brushing or washing, clinical reviewers reported that the standard of care no longer includes brushing or washing to collect specimens, as noted in the service descriptions. They also observed that the service description does not account for variation in the number of specimens collected by the physician, relative to the clinical indication for the procedure and their findings during the procedures; this concern has more to do with the HCPCS code descriptor than the intraservice description.

The service description for colonoscopy with polyp removal includes the use of a retrieval device, which clinical reviewers informed us they rarely use and, if so, only to remove unusually large polyps. Clinical reviewers disagreed with the service description stating that scopes are removed and re-inserted during the procedure with the removal of each polyp. Respondents informed us that the polyps are usually suctioned directly into a suction cup that also serves as the specimen sample collection cup and is directly sent to pathology after appropriate labeling. The service description does not mention collecting polyps in the suction cup, which clinical reviewers report has become the standard approach.

Postservice descriptions: Though they generally confirmed the accuracy of the postservice descriptions, clinical reviewers noted that they do not perform all activities listed. For example, clinical staff complete cytology and pathology forms and do the postprocedure specimen verification and documentation. Additionally, anesthesiologists and clinical staff, not the physician performing the procedure, routinely assess the patient for "suitability to be discharged from recovery suite."

Intraservice time estimates: Our respondents' intraservice time estimates are presented in table 9 (in the body of the report). For all five services, clinical reviewers estimated almost the same time as reflected in the PFS. There was consensus, however, that in cases where the physicians have to remove an unusual number of polyps or take additional biopsies, more time is taken—almost one to two minutes per additional polyp or additional biopsy specimen.

Ophthalmology

Summary: We reviewed four services: eye (intravitreal) drug injection (HCPCS code 67028), cataract surgery using intraocular lens, 1 stage (HCPCS code 66984), computer ophthalmic diagnostic imaging of the optic nerve (HCPCS code 92133), and computer ophthalmic diagnostic imaging of the retina (HCPCS code 92134). We discussed pre-, intra-, and postservice description accuracy for all four codes with clinical reviewers. They reported that for the two OCT imaging services (HCPCS codes 92133 and 92134) and the eye injection (HCPCS code 67028), the pre- and intraservice tasks are typically performed during a concurrent office visit, introducing the possibility that time spent during the office visit is inappropriately attributed to the imaging and procedure codes.

Vignettes: Clinical reviewers found the vignettes to represent typical patient presentations.

Preservice descriptions: Clinical reviewers found preservice descriptions to be accurate but made a few observations. For the preservice descriptions of OCT imaging services (HCPCS codes 92133 and 92134) and the eye injection (HCPCS code 67028), respondents noted that reviewing patient records and indications for the test, as mentioned in the service description, can be considered part of the concurrent office visit. They emphasized that there is a range of cognitive activity associated with the decision to order and interpret the test and the decision of whether to inject; this cognitive activity is reasonably attributable either to the office visit or to the service under consideration.

For eye injections, respondents reported that the five minutes of preservice "scrub, dress, and wait time" noted in the PFS does not align with current physician practices because the procedure does not require scrubbing and dressing.

Intraservice descriptions: For the most part, respondents found intraservice descriptions to be accurate but made a few observations. For eye injections, clinical reviewers reported no inaccuracies, but one clinical reviewer observed that a commonly used drug, Avastin, comes preloaded and ready to use, thus saving the minute or two that it would take to draw the drug into the syringe—an activity listed as an intraservice task in the service description. The same respondent said he performs the procedure often and described performing it for three patients in sequence in three different rooms because it takes about five minutes for the anesthetic to take effect; to improve efficiency, he multitasks by attending to other patients in the interim.

Clinical reviewers found the service description for cataract surgery to be accurate. Three respondents pointed out that a peripheral iridectomy, mentioned in the service description, is rarely performed. One said it would be "inappropriate" to do a peripheral iridectomy, since cataract surgery

itself is a substitute solution for it. The clinical reviewers also agreed that the described positioning of the soft patch and a rigid shield on the operative eye is done by others. In assessing preservice descriptions, respondents mentioned computerized OCT imaging (HCPCS codes 92133 and 92134) preservice tasks being performed concurrently with an office visit. That overlap holds true for OCT imaging intraservice tasks as well. Clinical reviewers told us they not only interpret the OCT images during the office visit, they also discuss the results with the patient during the same concurrent office visit while also discussing their overall status. Additionally, one respondent noted that sometimes, along with OCT imaging, other tests such as gonioscopy (a baseline visual field test) and Heidelberg retinal tomography are also done and interpreted simultaneously during the same concurrent office visit, indicating a range of interrelated judgments of which interpreting the OCT is but one component.

Though most clinical reviewers found the time and work taken for both OCT imaging services to be similar, two clinical reviewers noted that OCT of the retina often involves reviewing more images over a period of time intervals, suggesting that more time would be needed to interpret and review OCT of the retina than to interpret and review OCT of the optic nerve. The PFS gives both OCT imaging services an intraservice time of 10 minutes.

Postservice descriptions: Clinical reviewers did not report any inaccuracies in the postservice descriptions, although they thought that for eye injection and cataract surgery, some activities are now routinely performed by staff. The service description for eye injections assumes that the physician instructs the patient in postoperative care with topical medications, reviews the symptoms of potential complications, and completes the operative note. But three clinical reviewers informed us that these activities were performed by clinical staff. Similarly, for cataract surgery, they reported that they meet with the patient and the patient's family to discuss the procedure. However, clinical staff provide some other described postservice tasks, including giving patients post-op care instructions.

While the service descriptions for the two OCT imaging services did not list any postservice tasks, respondents agreed that once they interpret the results, they discuss them with the patient in detail—but again, they do this as part of the office visit, complicating attribution to one or the other service.

Intraservice time estimates: Our respondents' intraservice time estimates are presented in table 9 (in the body of the report). Broadly, the clinical reviewers' intraservice time estimates more or less matched the PFS intraservice times. Their major observation was the potential for conflation of OCT imaging intraservice time estimates with those activities provided in the concurrent office visit.

There was some variation in the intraservice times reported by clinical reviewers for cataract surgery. Our observations have a median value of 18 minutes, compared to the PFS time of 21 minutes,

but two respondents reported an intraservice time of 8 and 28 minutes respectively. The respondent reporting 8 minutes explained that he is faster than his colleagues because of his experience and efficiency. He estimated that his colleagues take 12 to 14 minutes or 15 to 17 minutes.

Role of technology: Two clinical reviewers mentioned that recent generations of OCT imaging machines have improved the reliability of computer-generated quantitate measurements and have also made it easier and quicker to review images across intervals to aid in the decision about treatment effectiveness.

Orthopedics

Summary: We studied four hip and/or knee surgical procedures: total hip arthroplasty (HCPCS code 27130), total knee arthroplasty (HCPCS code 27447), treatment of thigh fracture (HCPCS code 27236), and knee arthroscopy (HCPCS code 29881). We asked reviewers to comment on pre- and postservice description accuracy for total hip arthroplasty (HCPCS code 27130) and knee arthroscopy (HCPCS code 27236), and intraservice description accuracy for all codes. The reviewers estimated taking about the same intraservice time for these procedures as assumed in the PFS, but all agreed that the higher the patient's body mass index (BMI), the longer a procedure takes, with a "tipping point" to significantly increased time and work at a BMI of about 40. Furthermore, most of the tasks listed in the preservice description usually take place at a prior office visit days in advance, not on the day of the procedure.

Vignettes: Clinical reviewers agreed that the vignettes represent typical cases, but they observed that the BMI of the patient significantly affects the time required to do these procedures. With the exception of the knee arthroscopy vignette, which does not include the patient's BMI, the remaining three PFS vignettes indicated that the patient's BMI is "greater than 30." However, the respondents seemed to agree that a BMI of 40 or greater is the tipping point for substantially more time and effort, typically adding 20 to 30 minutes to the procedure.

Preservice descriptions: All clinical reviewers agreed that the preservice tasks in the service description occur mostly at prior office visits, thus rendering the descriptions inaccurate. The preservice tasks are performed at an office visit that occurs a few days prior to a knee arthroscopy procedure or are done during a couple of office visits spread over a couple of months prior to a major procedure such as total hip arthroplasty. Clinical reviewers explained that at these prior office visits, they review the preadmission imaging results, update the history and physical examination in the patient's chart, and review informed consent with the patient. One clinical reviewer who operates at an academic medical

center explained that they send patients to dedicated patient education units where they receive all relevant information a few days prior to the procedure.

Clinical reviewers said that on the day of the procedure, they typically review the patient's medical records and imaging and other lab results. They also meet with the patient in the preoperative area to identify and mark the correct site of the procedure, as stated in the service description. Respondents reported that they work regularly with clinical staff who verify that all required instruments and supplies are available for the procedures; the service description assumes physicians carry out these tasks.

Respondents noted an inaccuracy in the description of total hip arthroplasty. Since they no longer typically use intraoperative cell savers, they do not have to ensure the availability of equipment necessary for intraoperative cell savers, per the service description; instead, they use tranexamic acid to prevent blood loss during the surgery. The preservice description for total hip arthroplasty also incorrectly includes placing a tourniquet on the proximal thigh; all clinical reviewers reported doing that for total knee arthroplasties but not for hip arthroplasties.

Intraservice descriptions: Clinical reviewers agreed that the intraservice descriptions were mostly accurate but raised a few minor concerns. For total hip arthroplasties, they agreed that they use a C-arm to take an X-ray inside the operating room during the procedure, as mentioned in the service description, but they do so only if the procedure is performed through an anterior approach, which is not always the case.[18]

Three respondents noted that the service descriptions do not include a commonly performed activity—that is, using intra-articular blocks before completing total hip arthroplasty and thigh fracture treatment to provide postoperative pain relief.

Postservice descriptions: Clinical reviewers deemed the service descriptions accurate and agreed that they capture most of the postoperative tasks, but they indicated that the operating physician typically does not move the patient to the recovery area or initiate patient monitoring, as noted in the service description. The service description for knee arthroscopy mentions dictating an operative report and then writing procedure notes in the patient chart. Respondents informed us that these activities are not different and are both accomplished through a single EHR entry.

Intraservice time estimates: Our respondents' intraservice time estimates are presented in table 9 (in the body of the report). Reviewers reported times similar to those in the PFS for all four services. Though the estimates for the typical cases in the vignettes were close to the PFS times, the general

consensus was that the time taken for all four procedures can vary according to the BMI of the patient, with 20 to 30 minutes added for patients with BMI greater than 40 and incrementally more time added for increases in BMI above that threshold.

Role of personnel: Respondents reported receiving assistance from PAs during pre-, intra-, and postservice activities. One respondent reported being assisted by a PA for the last 21 years even during the intraservice period. Another respondent reported that on many occasions, the PA helps close the incision and performs other immediate postoperative tasks such as completing notes that the physician will later review. As discussed earlier, PA activities constitute physician work.

Radiology

Summary: We studied four services: brain MRI with and without contrast (HCPCS code 70553), frontal and lateral chest X-ray (HCPCS code 71020), spinal canal MRI without contrast (HCPCS code 72148), and thorax CT without contrast (HCPCS code 71250). We discussed pres- and postservice description accuracy for spinal canal MRI without contrast and thorax CT without contrast, and intraservice description accuracy for all codes. Reviewers noted that preparing case-by-case protocols is no longer part of physician preservice activities, and most radiologists concurrently review prior studies and interpret the current images as intraservice tasks rather than as preservice tasks, per the service descriptions. Both findings suggest that many of the preservice tasks are not performed according to the service descriptions.

Vignettes: Respondents generally confirmed vignette representativeness but raised concerns about vignettes for two codes—one that presented a patient with a known history of cancer (brain MRI, HCPCS code 70553), and one that presented a patient with pleural and pulmonary metastases (chest X-ray, HCPCS code 71020). The medical conditions of these hypothetical patients suggest that more than the typical time would be needed to rule out new metastases or to track growth of cancer. While not unusual, these presentations were atypical and would require more than the defined intraservice time.

Preservice descriptions: With few exceptions, the respondents said that they do not perform the noted preservice tasks, either because they are performed by the imaging technician or because they are routinely part of the intraservice period.

Respondents reported that radiology technicians now use established online protocols for specific imaging studies; issues arise only rarely, at which time the radiologist gets involved. Radiologists

develop and refine these protocols periodically—not with every patient, as the service description assumes. Only two of our five respondents assign these predesigned protocols to the patients the night before the procedure. The other three respondents reported that a technician reviews the patient's medical records and the referring physician's request, then performs the study based on the predesigned protocol for the given clinical scenario. The protocols are standardized, but even the two clinical reviewers who assign these predesigned protocols to each patient reported providing customized protocols in about 20 percent of cases.

Radiologists routinely review prior imaging studies, sometimes exhaustively but almost always as an integral part of intraservice work related to the study they are currently interpreting. Usually the prior studies are available in the PACS. Obtaining prior studies from other facilities can take time and is commonly handled by available assistants, such as film librarians and couriers. Only one clinical reviewer reported going into the archive a few times a day to obtain old studies during protocol determination—as the service description states—but this is rarely performed in practice and is not viewed as typical.

Respondents were divided over how often they thought radiologists review the appropriateness of the requested study and engage referring physicians to review and possibly modify imaging study requests (as indicated in the service description). One respondent noted that when a requested study is deemed inappropriate, it takes anywhere from three minutes to ten or more minutes to communicate with the referring physician and agree on an alternative.

Intraservice descriptions: Clinical reviewers observed that for three of the service codes (HCPCS codes 70553, 72148, and 71250), physician supervision of the technician work (noted in the service descriptions) does not occur. Respondents reported that their activities commonly begin when they are presented with a completed study on the PACS; this is when the respondents considered intraservice tasks to start. All of the respondents agreed with the service descriptions, except, as mentioned above, that they simultaneously review prior studies as intraservice tasks, not preservice tasks.

Postservice descriptions: With the widespread use of EHRs, which push out interpretations to ordering physicians, respondents agreed that they only reach out to such physicians in situations that necessitate immediate contact, such as when reading imaging studies ordered by the emergency room or when there are significant findings (e.g., an unexpected lung nodule). Personal contact with referring physicians is not typical as part of postservice work.

Role of technology: Four of the five respondents have been in practice since before the introduction of PACS, and all have had some exposure to interpreting images using film, the older method. They agreed

that PACS increased their ability to review images quickly. One respondent observed, and others agreed, that the improved image quality in PACS reduces per-image interpretation time and makes it easier to mentally construct three-dimensional pictures. However, some respondents noted that PACS substantially increases the number of images per study and that higher imaging quality increases the probability of making significant findings; both of these impacts serve to increase total time.

There was disagreement about the net effect on total time. Three clinical reviewers said that the various factors balanced out and that total time was about the same with PACS as before; one reviewer said total time was slightly lower with PACS, and another said it was slightly higher.

Intraservice time estimates: Our respondents' intraservice time estimates are presented in table 9 (in the body of the report). Most of their time estimates accorded with those in the PFS. Only one of the five respondents consistently offered lower time estimates for all codes. The clinical reviewers reported that the number and extent of significant findings, the nature of the underlying clinical condition (e.g., metastatic cancer), the number of prior studies to review, the age of the patient, or other factors are important considerations leading to variation in intraservice time. Although respondents thought two vignettes were atypical, they also thought that the typical interpretation takes longer because vignettes do not reflect that patients tend to be "older and sicker."

Urology

Summary: We studied five services: cystoscopy (HCPCS code 52000), cystoscopy with fulguration (HCPCS code 52224), transurethral resection of prostate (HCPCS code 52601), prostate biopsy (HCPCS code 55700), and prostatectomy (HCPCS code 55866). We discussed pres- and postservice description accuracy for prostate biopsy (HCPCS code 55700) and prostatectomy (HCPCS code 55866), and intraservice description accuracy for all codes. Reviewers pointed out that two vignettes presented atypical patient characteristics that could lead physicians to underestimate or overestimate the time required for the procedures. We also learned that for prostate biopsy (HCPCS code 55700), the intraservice description did not describe the role of ultrasound imaging during the procedure, perhaps underestimating the time and work involved in intraservice activities.

Vignettes: Respondents noted that vignettes were fairly representative, with two exceptions. Reviewers questioned the vignette for prostatectomy, which featured a 48-year-old male patient. The consensus was that the typical patient receiving this procedure would be older, likely in the Medicare age group. Because a younger patient would typically be healthier, with fewer comorbidities, this

vignette would likely result in atypically short time and work estimates, especially for preservice and postservice tasks.

Some respondents raised concerns about the vignette for TURP, which presents a patient with a prostate measuring 76 cubic centimeters. Respondents thought that this size was atypically large and, since the time and work of surgery depends on the amount of tissue removed, would produce exaggerated intraservice time and work estimates from physicians.

The vignette for cystourethroscopy with fulguration (including cryosurgery or laser surgery) created confusion among the respondents. The vignette presented a patient with carcinoma of the bladder who had undergone routine prior surveillance cystourethroscopy and had a number of lesions to be fulgurated and biopsied in this second cystourethroscopy. Respondents were unsure if this would typically be a first cystoscopy, rather than a follow-up, and expressed differing views about whether the intraservice description was correct, based on different assumptions about the patient's clinical problem. The vignette suggested a procedure that is usually performed in an office setting, but some reviewers thought the intraservice description for this service referred to a procedure usually done in an operating room.

Preservice descriptions: In general, respondents thought that the preservice descriptions for the office-based procedures were accurate but included tasks in the service description that are typically performed during a prior office visit. Such tasks include discussing the recommended procedure, obtaining consent, and performing the history, physical exam, and any needed laboratory tests.

Additionally, for office-based prostate biopsy, clinical reviewers report that some activities mentioned in the RUC preservice description are usually done by clinical staff. These activities include checking the schedule for the following day, making sure necessary instruments and personnel will be available for the procedure, making sure the ultrasound machine is available and working, confirming with scheduling staff that the patient was notified (now typically an automated procedure), confirming that the patient has taken enema and preoperative antibiotics, confirming that the patient is off of anticoagulants and/or nonsteroidal anti-inflammatory medications, and obtaining informed consent or verifying that informed consent was obtained. Respondents said they usually review the procedure and postoperative recovery with the patient and the patient's family at a prior office visit, not as part of preservice activities on the day of the procedure, as the service description assumes. The preservice tasks also include confirming that necessary imaging studies are available for review at the time of the planned procedure, which clinical reviewers felt is not common practice. Also, one clinical reviewer

reported ordering MRIs regularly, while another clinical reviewer mentioned doing so only for about 20 percent of her patients.

Similarly, many of the preservice tasks for prostatectomy are performed at a prior office visit. Respondents also noted that a few activities were missing from its preservice description, including prepping the patient and performing the surgeon's preoperative scrub and gowning.

Intraservice descriptions: For the most part, the service descriptions were accurate. According to the respondents we interviewed, intraservice tasks sometimes varied with workflow within the practice. However, respondents pointed out that the description for prostate biopsy omitted tasks associated with ultrasound guidance during the biopsy, a significant element in intraservice time and work.

Respondents generally agreed with the RUC intraservice description of prostate biopsy, but they noted that the description does not list using ultrasound imaging or guidance as part of intraservice activities. Ultrasound is mentioned in the RUC preservice description (in reference to the physician assuring the ultrasound machine is present and working) but is missing from the intraservice description. One clinical reviewer said that performing the prostate ultrasound takes a considerable amount of time and is performed along with an ultrasound technician. Respondents reported that in some cases they also bill for ultrasound interpretation, raising concerns about allocation of time and work to two concurrently performed codes. This respondent found ultrasound imaging to be helpful in looking for any areas suggestive of cancer and for measuring the prostate.

Additionally, while the service description indicates taking 12 biopsy samples, two respondents agreed that in rare cases, especially for patients who are on different protocols (i.e., patients on active surveillance or patients with imaging findings suggestive of cancer), more samples may be required, slightly increasing the work usually done for this procedure. The service description includes placing biopsy samples in labeled containers and applying rectal pressure, steps which respondents agreed were performed by clinical staff. Some respondents suggested that improvements in imaging technology have made the activities more intensive. Practices are using MRI technology more frequently, and even routinely, to identify possibly abnormal regions of the prostate. Reviewing the MRI and matching it to the ultrasound imaging takes additional time but, again, raises issues of allocation of time and work to different codes.

Respondents deemed the intraservice description for cystoscopy fairly accurate, but they noted that assessing bladder capacity, compliance, and sensation typically would not be done, contrary to the service description. Some clinical reviewers also reported having staff inject anesthetic jelly, apply the

penile clamp, and assemble the endoscopic equipment. Respondents reported that the common practice of taking urine samples for cytology was not included in the service description.

For fulguration, respondents' main observation was that the vignette describes a case which may or may not be performed in the operating room. Respondents observed that, if the procedure was performed in the operating room, a rigid (as opposed to flexible) scope would be used. Moreover, most respondents use a Bugbee cautery, not a laser fiber, for the lesions. Blue light cystoscopy, in addition to white light cystoscopy, is now sometimes performed; this takes additional time and may lead to more biopsies and fulgurations. Respondents split over whether rectal swab or urine cytologies were routine.

Postservice descriptions: For prostate biopsies and prostatectomies, the service description assumes that physicians transfer patients from the operating table or procedure room to the recovery area, but respondents informed us that clinical staff routinely does this. The prostate biopsy, performed nearly 70 percent of the time in a physician's office (according to Medicare claims data), does not involve transferring a patient to a postoperative stretcher and recovery area, as the RUC postservice description states. Rather, as one clinical reviewer summarized, when the procedure ends, patients typically "get up, get dressed, and leave."

Intraservice time estimates: Our respondents' intraservice time estimates are presented in table 9 (in the body of the report). Most of their intraservice time estimates were similar to those in the PFS and to each other, with a few exceptions. For prostatectomy, respondents noted that docking the robot can sometimes take a significant amount of time, depending on the staff's experience. Respondents also noted that some surgeons work slower than others. For cystoscopy with fulguration, time estimates varied; this was related to the confusion over whether the procedure is done in an office setting or in an operating room. Additionally, time estimates varied because of differences in the number of lesions that needed to be fulgurated or biopsied, and if blue light cystoscopy was performed in addition to white light cystoscopy. For TURP, the respondents' time estimates varied because of differing physician practice settings and differing assumptions about the typical size of the prostate.

Appendix C. Challenges of Empirical Time Data Collection

Even before we got into the field, potential and participating sites brought up a number of unexpected issues related to the data collection process. These issues ranged from concerns about union work rules and the provider's IRB requirements to low service volumes for study services and data system limitations. In response to some specific issues, we modified our data collection approach. For some services in some sites, our project staff were not allowed to directly observe patient care areas. Instead we trained site staff to do the observation, while our staff were on-site to oversee their work. We developed site-specific data collection plans to accommodate each site's data systems and clinical organization and to respond to IRB and other concerns. We developed data collection protocols for both direct observation and electronic time data. These protocols were used both for training and for field reference. The interim report describes many of these issues and the final protocols in detail (Zuckerman et al. 2014).

While working with sites to collect data, we encountered a number of other obstacles to our original project plan. These considerations affected how we collected data at the three sites and may inform future efforts to collect this type of data. Some of these challenges could affect the reliability of the study findings as they relate to the time measures for specific HCPCS codes. This appendix describes some of these challenges.

Even after staff agreed to participate, data collection required significant buy-in across the health care organization. Data collection engaged the entire organization, and each step in the process typically involved different people. These steps included agreeing to a draft scope of work, reviewing and approving the subcontract with the Urban Institute, getting IRB approvals, conferring with the IT department to gain an adequate understanding of its systems and capabilities, and working out a process for direct observation, which included assigning internal staff to the tasks. Engaged site staff generally had competing demands on their time and often found it difficult to give these data collection efforts high priority. This created delays in the data collection process and made it difficult for project staff to efficiently plan an approach to completing the various tasks.

The three sites responded to data collection challenges in different ways. One site opted to forgo the potentially lengthy IRB process that would have been required for our outside observers and instead chose to use their own staff for direct observation data collection. The staff lead was involved in

every aspect of the study from start to finish, including day-to-day interaction with department observers. When a potential issue or challenge arose, the internal team proposed a remediation strategy to address it. This was the only site that succeeded in meeting volume targets, but volume was still low because this site only did direct observation. At another site, data collection was much more difficult. Specific approvals were needed for any time staff spent on the study (including for administrative tasks), and substitute staff were needed to fill in for clinical staff when they were absent from their regularly scheduled tasks to participate in this study.

Service volumes were low for many HCPCS codes at participating sites. As discussed in our interim report, one criterion used to select the study's HCPCS codes was high frequency among Medicare FFS beneficiaries (Zuckerman et al. 2014). Each study site provided estimates of the annual volume for each of the 117 HCPCS codes targeted in our study. Study sites also identified which HCPCS codes would have electronic time data. A review of typical weekly service volumes for our selected codes across the three sites showed lower than expected volumes, which required us to adjust the frequency of direct observation events per code. We identified HCPCS codes with weekly volumes of 10 or more, if there was no electronic time data, or 3 or more, if electronic time data was available. This produced the list of HCPCS codes and volumes targeted for direct observation. However, once we were in the field, we concluded that achieving even these scaled-back targeted volumes within a five-day observation period was not realistic.

Several factors affected our ability to observe targeted volumes of the study's HCPCS codes. First, it was often difficult to predict what service would actually be performed and what final HCPCS code would be assigned to a bill. We found that scheduled procedures could be changed due to timing conflicts or clinical decisions (e.g., once surgery started, the surgeon may realize that a different procedure was necessary or that the anticipated procedure was unnecessary). Second, we could not always schedule data collection to coincide with periods of highest expected volumes. Data collection dates were affected by factors such as the department's ability to host observers, availability of internal observers, and time of year during which data collection took place (e.g., the desire to perform data collection prior to December). Third, the need to allocate direct observation staff across multiple surgical suites, each being used to perform multiple procedures in a given day, limited our ability to collect direct observation data for as many procedures as we had anticipated. The site that was most successful in meeting target volumes used their own staff to perform direct observations over a period of several weeks.

Identifying specific HCPCS codes for scheduled procedures presents a significant challenge for direct observation. Discussion with clinical leaders within each of the study sites confirmed that for

some services, the scheduled procedure description will likely be the final performed procedure. However, for some services, such as screening colonoscopy, there is some uncertainty about whether or not the colonoscopy will include biopsy, polyp removal, or other findings that would alter the assigned HCPCS code. Thus, we were not always conducting direct observation for one of the targeted HCPCS codes, reducing the number of direct observations for the study.

We concluded that future efforts to collect data via direct observation should consider identifying physicians who can be observed in their daily routines as opposed to targeting specific HCPCS codes. Because of the difficulty of predicting the exact service (and HCPCS code) that is going to be provided during a scheduled visit or even a schedule procedure, many observed patient-physician encounters did not lead to the expected service and HCPCS code. Thus, we believe that recording and timing all of the services that a physician provides within a sufficiently long time period—perhaps a few days to a week—would allow observation of a reasonable sample of high-volume services.

For one case, ECG interpretation (HCPCS code 93010), we modified our approach to reflect the clinical reality of how the service is provided. We had a study team member sit and observe physicians as they read ECG reports for a number of patients at once. This physician-centric, rather than patient- or service-centric, approach worked well in this instance, yielding time estimates for hundreds of cases in a very efficient way.

Observing or collecting time data for individual HCPCS codes can be complicated. As we collected time data using either direct observation or time stamps from EHRs, it became clear that multiple HCPCS codes are often provided and billed for during a single patient-physician contact. The need to adjust payments to account for multiple services provided in the same encounter has been well recognized in Medicare physician payment systems for many years. However, making adjustments for time is more complicated because it is often hard to know when one HCPCS code ends and another begins. For example, some ophthalmological diagnostic tests that are billable in addition to a visit are so fully incorporated into the visit interaction that the service-specific elements are hard to observe, much less time accurately. Situations in which we observed more than one HCPCS code impeded our ability to analyze the data in this study.

Select high volume procedures are heterogeneous. During site visits, clinical leaders expressed concern that time estimates derived for a few of the selected HCPCS codes may not be representative of the time necessary to perform all elements of the service because of extreme clinical heterogeneity. Two examples were cited: (1) Cystourethroscopy (HCPCS code 52000) could be a short uncomplicated follow-up procedure after treatment for cancer, taking very few minutes, or a lengthy initial diagnostic

procedure for a newly diagnosed cancer patient. (2) Level IV surgical pathology (HCPCS code 88305) can be performed on one or multiple tissue samples, and the elements can differ depending on the nature of the request and tissue source(s).

Available electronic data tends to be focused on intraservice time for surgical procedures, including endoscopies. In discussions with health IT staff at the two sites that had EHR data, we confirmed that the strongest electronic data pertain to surgical procedures. This finding is consistent with the MedPAC report, which concluded that service time, especially intraservice time, is most available for major surgical services (Braun and McCall 2011). We also confirmed that pre- and postservice times for most of the HCPCS codes are not well captured in electronic data and are difficult to observe. Electronic time data are often unavailable, most notably for office-based procedures. While time stamp data are generally available for imaging services, there are substantial interruptions during the intraservice period of interpretation and reporting, which reflect pre- and postservice tasks for other services to other patients (i.e., consulting with a technician on setup or consulting with the ordering physician on an interpretation of prior studies). These findings convinced us to modify the direct observation data collection protocol and tool.

Notes

1. A detailed description of the fee schedule's background and related policies can be found in Centers for Medicare & Medicaid Services, "Final Rule on Revisions to Payment Policies under the Physician Fee Schedule, Clinical Laboratory Fee Schedule, and Other Revisions to Part B for CY 2014," 78 Fed. Reg. 74230 (Dec. 10, 2013).
2. Participants in the clinical review interviews are referred to as "clinical reviewers" or "respondents" in this report.
3. The Protecting Access to Medicare Act of 2014 includes an expanded list of categories of services that may be misvalued.
4. For audiometry testing (HCPCS code 92557), observers recorded intraservice tasks as being done by providers of type "other." Since we could not discern whether this was a nonphysician provider (such as an audiologist) or clinical staff, we dropped the service from the analysis.
5. For the two angiogram codes, the reported distribution of utilization in the RUC database seemed clinically implausible, with "physician's office" reported as the dominant setting. According to CMS utilization data, these codes are paid in a facility setting over 99 percent of the time, using the professional-only (26) modifier. Thus, we classified these codes in the outpatient type rather than the office-based type.
6. Of the six services, four were imaging and test interpretations, including two radiology (brain MRI with or without contrast, and spinal canal MRI without contrast) and two optical coherence tomography imaging services. We studied eye injections alongside cataract surgery (for which we had empirical data) and knee arthroscopies alongside three major hip and knee surgical procedures (for which we had empirical data).
7. The clinical director of this project, Robert Berenson, was the lead developer of the widely used Berenson-Eggers Type of service classification system and used similar principles for creating service groupings for this analysis.
8. All PFS data used in this report were drawn from Centers for Medicare & Medicaid Services, "Final Rule on Revisions to Payment Policies under the Physician Fee Schedule and Other Revisions to Part B for CY 2016," 80 Fed. Reg. 70886 (Nov. 16, 2015), https://www.federalregister.gov/articles/2015/11/16/2015-28005/medicare-program-revisions-to-payment-policies-under-the-physician-fee-schedule-and-other-revisions.
9. This method is explained in detail in Wynn et al., *Development of a Model for the Validation of Work Relative Value Units for the Medicare Physician Fee Schedule* (Santa Monica, CA: RAND, 2015).
10. Wynn et al. also consider these three alternatives for intraservice work in *Validation of Work Relative Value Units* (110–112), labeling them "Increased IWPUT," "No Change in Mean IWPUT Value," and "Blend."
11. Two of the study's HCPCS codes, 93010 and 88305, have such large differences in implied intensity under the two measures that we have had to omit them from analyses that are overly influenced by outliers. HCPCS code 93010 was left out of figure 1 because it would be difficult to see most of the cases if the vertical scale was extended to 1.5 to accommodate this code.
12. HCPCS codes 93010 and 88305 were included in the calculation of the median because magnitude does not exert undue influence on the median.
13. This overall shift in intraservice intensity would raise potential issues within the building block method, however. Our implied intraservice work values are based on specific numeric intensity assumptions for preservice and postservice work, which have presumably been based on their relationship to the level of intraservice intensity. If intraservice intensity is generally 50 percent higher than previously thought, then it would be important to reassess the specific intensity values assumed for pre- and postservice work. If these

values should be increased as well, then estimates of intraservice work would decline. If pre- and postservice intensities are *not* changed, then even though intraservice work is changed proportionately, the effect on total work would not be proportionate.

14. Variation is measured as the adjusted R-squared of an OLS regression of study-based intensity as a function of PFS intensity. Outlier HCPCS codes 93010 and 88305 were omitted from this calculation.

15. We conducted a sensitivity analysis that showed that excluding interpretation services *increases* the share of variation in study intraservice intensity accounted for by PFS intensity. This involved re-estimating the regression described in note 14, omitting each type of service category one at a time. The adjusted R-squared is 0.3856 based on all 58 cases; it increases to 0.5125 when we omit the 10 imaging and other test interpretations and falls to 0.2903 when we omit the 19 OPD/ASC services. We omitted HCPCS codes 93010 and 88305 from all of these analyses.

16. For this example, intensities were rounded to the nearest 0.01.

17. Reviewers identified a number of other discrepancies. For cataract surgeries (HCPCS code 66984), physicians consider the described peripheral iridectomy inappropriate for the typical case. For endoscopy (HCPCS code 43235) and colonoscopy (HCPCS code 45378), physicians do not typically perform brushing and washing to collect specimens. Instead, they take biopsies and use a different code for those biopsies. For colonoscopy with polyp removal (HCPCS code 45385), the service description notes the use of a retrieval device and the need to withdraw the scope with the removal of every polyp. Physicians report that this is not how the procedure is typically performed; now they extract polyps by the snare technique, suctioning them into a collection cup directly—a time-saving step missing from the current description. Physicians say that retrieval devices are used rarely, if at all, and only to extract large polyps.

18. Respondents reported using different approaches based on their experience and the requirements of the case. We did not collect further information on the frequency or preference of either (anterior or posterior) approach.

References

Braun, Peter, and Nancy McCall. 2011. *Improving the Accuracy of Time in the Medicare Physician Fee Schedule: Feasibility of Using Extant Data and of Collecting Primary Data*. Washington, DC: MedPAC.

McCall, Nancy, Jerry Cromwell, and Peter Braun. 2006. "Validation of Physician Survey Estimates of Surgical Time Using Operating Room Logs." *Medical Care Research and Review* 63 (6): 1–14.

Wynn, Barbara O., Lane F. Burgette, Andrew W. Mulcahy, Edward N. Okeke, Ian Brantley, Neema Iyer, Teague Ruder, and Ateev Mehrotra. 2015. Development of a Model for the Validation of Work Relative Value Units for the Medicare Physician Fee Schedule. Santa Monica, CA: RAND Corporation.

Zuckerman, Stephen, Robert Berenson, Katie Merrell, Tyler Oberlander, Nancy McCall, Rebecca Lewis, Sue Mitchell, and Madhu Shrestha. 2014. *Development of a Model for the Valuation of Work Relative Value Units: Objective Service Time Task Status Report*. Washington, DC: Urban Institute.

About the Authors

Stephen Zuckerman is a senior fellow and codirector of the Health Policy Center at the Urban Institute. He has studied health economics and health policy for almost 30 years and is a national expert on Medicare and Medicaid physician payment, including how payments affect enrollee access to care and the volume of services they receive. He is currently examining how payment and delivery system reforms can affect the availability of primary care services, and studying the implementation and impact of the Affordable Care Act.

Katie Merrell, formerly of Social & Scientific Systems, is a senior research scientist at Actuarial Research Corporation. Since joining the staff of the Physician Payment Review Commission in 1990, Merrell has done extensive work in the area of Medicare physician payment policy and other aspects of physician compensation, organization, and training. In addition to this research, Merrell has led studies of Medicare and Medicaid managed care, the Medicare drug benefit, and evaluations of several maternal and child health initiatives.

Robert Berenson joined Urban as an Institute fellow in 2003. He conducts research and provides policy analysis primarily on health care delivery issues, particularly related to Medicare payment policy, pricing power in commercial insurance markets, and new forms of health delivery based on reinvigorated primary care practices.

Susan Mitchell is a research health IT scientist with RTI International's Digital Health Policy and Standards Program. Her professional concentrations include health and information management systems, technologies, and standards.

Divvy Kant Upadhyay is a research associate in the Health Policy Center at the Urban Institute, where he focuses on qualitative research on health reform measures at the federal and state levels. His broad range of policy research interests include the Affordable Care Act; reforms in health care delivery, physician payment, and federal programs such as Medicare and Medicaid; health care pricing and markets, and antitrust laws; and issues in medicine related to primary care, family practice, and diagnosis errors.

Rebecca Lewis is a public health analyst with RTI International's Health Care Quality and Outcomes Program. Her professional concentrations include Medicare payment systems and implementation and evaluation of quality performance.

Statement of Independence

The Urban Institute strives to meet the highest standards of integrity and quality in its research and analyses and in the evidence-based policy recommendations offered by its researchers and experts. We believe that operating consistent with the values of independence, rigor, and transparency is essential to maintaining those standards. As an organization, the Urban Institute does not take positions on issues, but it does empower and support its experts in sharing their own evidence-based views and policy recommendations that have been shaped by scholarship. Funders do not determine our research findings or the insights and recommendations of our experts. Urban scholars and experts are expected to be objective and follow the evidence wherever it may lead.

2100 M Street NW
Washington, DC 20037

www.urban.org