September 9, 2026

The Honorable Mehmet Oz
Administrator
Centers for Medicare & Medicaid Services
Department of Health and Human Services
7500 Security Boulevard
Baltimore, Maryland 21244-1850

Via Electronic Submission: https://www.regulations.gov

Re: CMS-1848-P; CY 2027 Revisions to Payment Policies under the Physician Fee Schedule and Other Changes to Part B Payment Policies; Current Procedural Terminology Request for Information (July 16, 2026)

Dear Administrator Oz,

Thank you for the opportunity to comment on the CY 2027 proposed rule, and for the questions posed in the Request for Information at section II.H. I am Principal and founder of Savage Health Policy, LLC, a health policy consulting practice established in 2022. I have spent more than two decades in health policy and government relations, working on Medicare coverage and payment, quality measurement, and the approval and coverage of medical devices and prescription drugs. Before founding the firm I served as Senior Vice President of Policy and Advocacy at the Cancer Support Community, and before that as Vice President of Government Relations for the Society of Thoracic Surgeons. I write in response to Question 4 of the Request for Information, "What objective alternatives exist, or could be developed, to maintain a more objective process to the current AMA CPT and RUC committee processes? How would these alternatives support or inhibit innovation?"

Summary

The Resource-Based Relative Value System (RBRVS) rests on a principle that no part of the valuation process enforces: relativity. The statute defines physician work as time and intensity. Time is measured; intensity is not. Time alone now predicts 81 percent of the variation in work values, and no step in the process tests any value for consistency with the values around it.

That test can be built using CMS's published files and the survey record of the AMA/Specialty Society Relative Value Scale Update Committee (RUC). Each service receives a measured position based on the character of its work, and each value is compared with the values of the services nearest it. These comments propose a New RBRVS built on that foundation, one where consistency rules are stated in writing and

every value is tested against comparable work. Values that fail the test are corrected by rule, in public view, without anchor codes and without negotiation.

A schedule with stated rules to govern it changes what rulemaking can do. CMS could model a proposed policy, run it against the entire schedule, and publish the projected results before adopting it. Public comment would address the projected outcome of a stated policy rather than thousands of individual values whose combined effect no one can see.

The same standard resolves the Request for Information's innovation question. A new service can receive a defensible value from measured comparison with existing work, not from the strength of its constituency.

Any system will benefit from the best data available. Therefore, I support the empirical time data program CMS proposes. Empirical measurement addresses time, the observable half of the statutory definition. Intensity has no empirical source; it can be established only by comparing each service's value with the values of comparable work. I have developed a test that makes that comparison across the whole schedule, and these comments describe it.

Current Procedural Terminology (CPT) Request for Information

CMS has a long-held skepticism of the RBRVS that is expressed as distrust of the RUC. In this proposed rule, CMS noted nearly 20 years of concern from the Medicare Payment Advisory Commission (MedPAC) that CMS has "over-relied on specialty societies with a financial stake in the process," and has articulated a "longstanding concern expressed over the Federal reliance on a private organization with such an obvious conflict of interest as providing information on the time and resource requirements to conduct physician services when this information may influence their own payment."[1] Elsewhere in the same rule, CMS faults its own practice expense methodology for a method that "privileges the historic survey data over incorporation of more recent data about specific services and produces unpredictable and counterintuitive results that are not transparent to the public."[2] And at section II.D of the rule, CMS states that empirical data "may be a better input than limited survey data."[3]

CMS's diagnosis is that the valuation process rests on the estimates of interested parties, privileges historic survey data over current evidence, and produces results the public cannot check. My diagnosis is that the fee schedule does not enforce the principle it is named for: relativity. First, the schedule does not measure work as the statute defines it. Time is measured; intensity is not. Making matters worse, time has come to stand in for intensity system-wide. Second, nothing in the process tests any value against the schedule as a whole: relativity is asserted, never checked. This letter demonstrates that the missing test requires no committee that can see the whole schedule—only local rules and a measured position for every service. I have built much of that infrastructure from CMS's published files and the RUC's own survey record. The Analysis below documents this diagnosis; the Proposal that follows is my answer to Question 4.

Analysis

The relative value unit (RVU) consists of physician work, practice expense (PE), and malpractice. Section 1848(c)(1)(A) of the Social Security Act defines the work component as "the portion of the resources used in furnishing the service that reflects physician time and intensity in furnishing the service," a definition CMS restates in the preamble of the

---

[1] Medicare and Medicaid Programs; CY 2027 Payment Policies Under the Physician Fee Schedule and Other Changes to Part B Payment and Coverage Policies; Medicare Shared Savings Program Requirements; and Medicare Prescription Drug Inflation Rebate Program, Proposed Rule, 91 Fed. Reg. 43,842 (July 16, 2026) (CMS-1848-P) [hereinafter Proposed Rule]. Quotations at 91 Fed. Reg. 43,952.

[2] Proposed Rule, 91 Fed. Reg. at 43,850.

[3] Proposed Rule, 91 Fed. Reg. at 43,934 (section II.D).

proposed rule.[4] PE and malpractice may carry biases of their own; I do not address them here.

The definition's two quantities differ in kind.

- **Time** is observable and can be measured accurately. The RUC uses surveys and other methods to collect time data, and clinical registries capture time as a byproduct of documentation. CMS already names Veterans Health Administration data, the National Surgical Quality Improvement Program, the Society of Thoracic Surgeons National Database, and Merit-based Incentive Payment System data among the sources it will consider for "analyses of work time, work RVUs, or direct practice expense (PE) inputs."[5] The empirical infrastructure for the time half of the definition exists, and it is improving.
- **Intensity** cannot be counted. It appears in no claim, timestamp, or operative log. In 1988 William Hsiao and his Harvard colleagues, whose research built the original relative value scale, named its components alongside time: mental effort and judgment, technical skill and physical effort, and psychological stress.[6] Each can be established only by comparison.

Both halves of the definition are collected today by one instrument: the RUC survey. When a service comes up for valuation, the relevant specialty society surveys physicians who perform it, and each respondent supplies four things:

- Time estimates for the service—pre-service, intra-service (the procedure itself), and post-service.
- A reference service, chosen from a list the society provides, that the respondent judges most similar to the service being surveyed.
- Component ratings against that reference: whether the surveyed service requires more or less mental effort and judgment, more or less technical skill, more or less physical effort, and more or less psychological stress (the last including the risk of significant complications), each on a 5-point scale from much less to much more.
- A proposed work value for the service—the respondent's own estimate of the number. (The survey form, before and after the 2017 change described below, is reproduced at Appendix A.)

Of these four, the RUC debates the times and the proposed value, and CMS reviews them. The component ratings are not cited, although they are the only part of the survey that measures intensity as the statute defines it.

---

[4] Proposed Rule, 91 Fed. Reg. at 43,866; Social Security Act § 1848(c)(1)(A), 42 U.S.C. § 1395w-4(c)(1)(A).

[5] Proposed Rule, 91 Fed. Reg. at 43,929.

[6] Hsiao WC, Braun P, Dunn D, Becker ER. Estimating Physicians' Work for a Resource-Based Relative-Value Scale. *New England Journal of Medicine*. 1988;319(13):835–841.

Interestingly, the intensity part of that instrument was also made coarser in 2017. Until then the survey asked eight component questions in addition to an overall intensity judgment: three addressing mental effort and judgment, three addressing psychological stress, and one each for technical skill and physical effort. In April 2016 the RUC's Time and Intensity Workgroup recorded that the measures were "hard/time-consuming to interpret" and that "there is simply too much information to be able to review it all effectively." In the same passage it recorded a validity concern: responses tended "to most commonly indicate that the survey code is somewhat more intense than the reference code"—a leaning toward the reference, observed by the committee in its own data.[7] In January 2017 the Research Subcommittee approved collapsing "the three mental effort and judgment and three psychological stress intensity and complexity questions each into a single question."[8] The stated reasons were reviewer burden and response rate. The review burden that justified the coarsening is precisely the kind of processing burden modern analytical tools now eliminate.

Intensity is never recovered later in the valuation arithmetic. In 2011, in work commissioned by MedPAC, Braun, a co-author of the original RBRVS research, and McCall examined the RUC's Building Block Method, which assembles a recommended value from separately estimated components. They found that the method derives intraservice work intensity as a residual rather than measuring it, producing "anomalously low or even negative values"[9] of intensity. Intensity is collected by the survey, unused in the debate, and reconstructed as a leftover when it appears at all.

This time/intensity asymmetry is compounded because time often substitutes for intensity in the work calculation itself. For particular codes, CMS states, "we refine the work RVUs in direct proportion to the changes in the best information regarding the time resources involved in furnishing particular services."[10] Work then moves with time, and intensity is held constant by construction. By this mechanism a schedule defined as time and intensity comes to behave as a schedule about time. The statute does not direct that result.

CMS is not alone in using time as a fallback measure. In its most recent recommendations, the RUC routinely arrives at a work value by crosswalk, adopting the established value of a comparator service it selects. Intraservice time is the variable that selects the comparator,

---

[7] AMA/Specialty Society RVS Update Committee, April 2016 Meeting Minutes, Time and Intensity Workgroup, Intensity and Complexity (I/C) Measures, at 67–68.
[8] AMA/Specialty Society RVS Update Committee, January 11–14, 2017, Meeting Minutes, Research Subcommittee report, at 95.
[9] Braun P, McCall N. *Methodological Concerns with the Medicare RBRVS Payment System and Recommendations for Additional Study.* RTI International, prepared for the Medicare Payment Advisory Commission. December 2011.
[10] Proposed Rule, 91 Fed. Reg. at 43,866.

even though a survey was conducted in each case. Four examples from the February 2026 report alone:[11]

- Repair of a chronic traumatic diaphragmatic hernia via thoracotomy. The RUC "reviewed survey results from 32 trauma physicians and recommends a work RVU of 26.41 based on a direct crosswalk to CPT code 35601 Bypass graft, with other than vein; common carotid-ipsilateral internal carotid (work RVU = 26.41, 180 minutes intra-service time), which appropriately accounts for the physician work typically required to perform this service." A diaphragm repair is valued at a carotid bypass graft, and the two share an intraservice time of 180 minutes. The two further codes the RUC cites in support carry the same 180 minutes.
- Repair of a nontraumatic diaphragmatic hernia via thoracotomy, valued by direct crosswalk to a laparoscopic paraesophageal hernia repair. An open repair is valued at a laparoscopic one; both carry 180 minutes of intraservice time.
- The add-on code for implanting mesh, valued against iliac revascularization codes that share, in the RUC's stated basis, the same pair of quantities: "30 minutes intra-service time and a work RVU of 3.00."
- Injection of sclerosant, multiple incompetent veins other than telangiectasia, same leg, valued by direct crosswalk to repositioning a central venous catheter. Here the report states the mechanism outright: the surveyed "intra-service time decreased by 10 minutes compared to the existing time, while the intensity of the service has remained unchanged," and the crosswalk "more accurately reflects the reduced intra-service time without changing the intensity…"

These may be sound clinical judgments; experienced clinicians made each comparison. But time is the operative variable in every example. In all four, the comparator carries the same intraservice time as the surveyed service: 180, 180, 30, and 20 minutes respectively. The observable half of the work component is doing the work that intensity was meant to do. CMS's own published files show the same pattern schedule-wide, with no survey inputs and no model assumptions.

This means a service's work RVU can be predicted from its intraservice time. I found that time alone explains 81 percent of the variation in work values across the 6,415 services with a recorded intraservice time, 86 percent when both are measured in logarithms, and the strength of that relationship has not moved since 2015.

[11] AMA/Specialty Society RVS Update Committee, February 2026 Meeting Recommendations. Quotations verified against the published report.

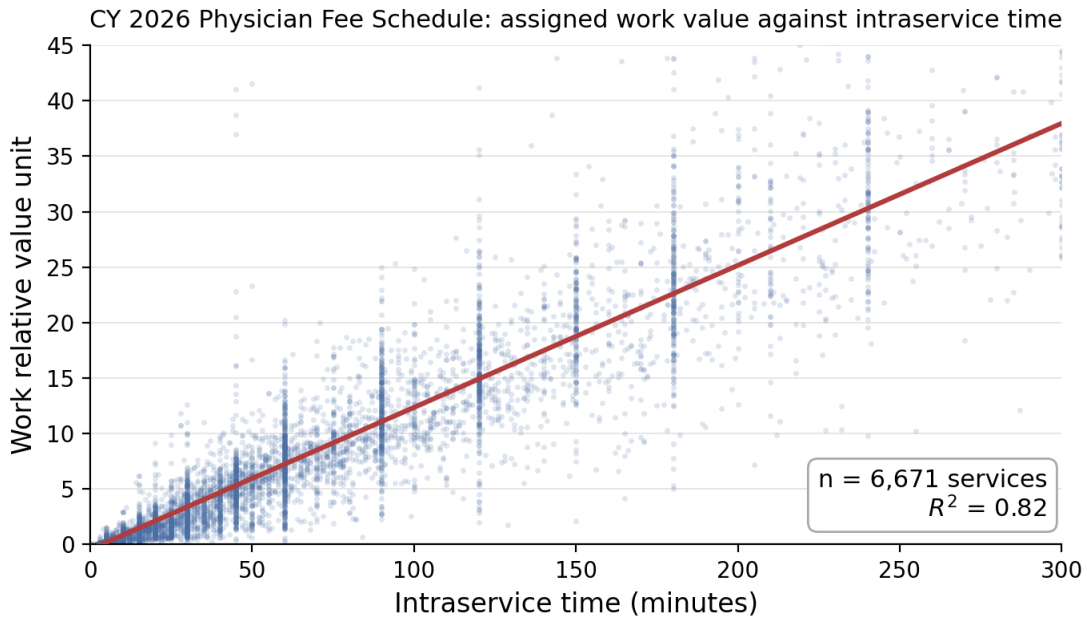


Figure 1: Work RVU vs intraservice time, CY 2026 Physician Fee Schedule
Figure 1. Each point is one service in the CY 2026 Physician Fee Schedule. Assigned work value tracks intraservice time closely and the relationship is linear across the full range. The vertical bands at 30, 60, 90, 120 and 180 minutes are services whose assigned times cluster on round numbers.

However, the schedule does not attribute a uniform value per minute. Across the 6,671 services in the CY 2026 schedule, value per intraservice minute varies by a factor of 3.7 between the 10th and 90th percentiles—the variation where intensity should be visible. It is not: on the 282 services whose intensity ratings are published in usable numeric form, adding measured intensity to time improves the prediction by about one tenth of one percentage point (in logarithms the fit moves from 0.855 to 0.856; in levels, from 0.922 to 0.923). Part of time's dominance is arithmetic: work is time multiplied by intensity, and time varies far more.

Arithmetic does not explain why measured intensity adds so little. And time's dominance is only half the problem. The other half is that nothing in the process holds any work value consistent with the rest of the schedule, despite the best efforts of the RUC and CMS.

The RBRVS does not ignore relativity. Surveys are built on comparisons to key reference services that each respondent selects from a society-provided list. RUC deliberation adheres to certain rules, all intended to enforce relativity, including guards against rank-order anomalies within a family, and crosswalks binding one service's value to another's. CMS "utilize[s] an incremental methodology in which [they] value a code based upon its incremental difference between another code and another family of codes,"[12] and CMS's

[12] Proposed Rule, 91 Fed. Reg. at 43,866.

criteria for identifying potentially misvalued services name "anomalies in relative values within a family of codes" and "an anomalous relationship between the code being proposed for review and other codes" as reasons for additional scrutiny.[13]

These are important principles, but none of them tests the system as a whole. Every route of review produces a list of services to examine, never a test a value passes or fails. And no step reconciles separate comparisons: one service is crosswalked to a second, a third to a fourth, and nothing asks whether the first and third now sit correctly against each other.

The scale of these unreconciled comparisons is knowable, because survey respondents declare their reference service on the public record. From those declarations I was able to reconstruct the network of comparisons in the RUC's recommendation records: the 2010 Five-Year Review and every meeting report from October 2014 through May 2026.[14] The network contains 2,225 services joined by 3,066 links, and it is not fragmented: 1,847 of them — 83 percent — form a single connected component, one continuous web, with the remainder scattered in small islands of no more than 29 services. The web shows how densely the record is cross-referenced. It does not by itself establish that distant services are comparable; that is what the measured positions are for.

The declared references also understate the comparison activity: the committee's own justifications routinely reach past the reference services the survey respondents chose. In a verified sample of 356 valuation recommendations drawn from five RUC reports spanning 2010 through 2025, 90% of the committee's value justifications relied on comparator codes selected outside the survey respondents' own key reference choice. The sample was verified by hand against the source recommendations; the hand check agreed with the coded output on 18 of 20. The RUC's own records show that the comparisons are already made, dense, cross-cutting, and the operative step in setting a value. No one has stated them as rules or checked them for consistency.

Two further observations complete the picture, and neither comes from any model.[15] First, across implementation years 2015 through 2026, CMS was more likely to reduce a recommended value when the service is high volume—not when it is merely expensive per use. That is a conditional association, not a causal claim. Second, the reductions fall

---

13 Proposed Rule, 91 Fed. Reg. at 43,930.

14 Author's analysis of declared key reference services in AMA/Specialty Society RVS Update Committee meeting recommendation records (the 2010 Five-Year Review; meeting reports October 2014–May 2026). Methodology, code, and analysis-ready data available to CMS staff on request. Each link is a survey respondent's declared key reference for a surveyed service, first or second choice; repeated declarations across years are counted once.

15 Author's analysis of RUC work-value recommendations and CMS final actions, implementation years 2015–2026, from CMS's published Physician Fee Schedule relative value files and RUC meeting recommendation records. Available to CMS staff on request.

unevenly: across 1,049 RUC work-value recommendations that CMS acted on for implementation years 2015 through 2026, CMS finalized a work value below the RUC's recommendation for 139 of 511 recommendations in the surgical code range (a range defined by CPT numbering, not a clinical classification), or 27.2%, against 69 of the remaining 538 recommendations in all other code ranges, or 12.8%—roughly twice the rate. The unit here is a recommendation rather than a distinct code, and the comparison sets aside the codes reduced by the CY2026 efficiency adjustment; the difference holds under either treatment. These are observations: they do not explain one another, and they say nothing about intent. What they show is an agency already adjusting the schedule in patterned ways, with no published standard against which anyone—including CMS—can check the pattern.

Proposal: A New RBRVS

CMS asks: "(4) What objective alternatives exist, or could be developed, to maintain a more objective process [than] the current AMA CPT and RUC committee processes? How would these alternatives support or inhibit innovation?"[16] Everything in this section is my answer to that question. It is a proposal, not a description of the current system. Where a piece of the proposal has been built and run, it is included here.

I propose a New RBRVS with consistency rules stated in writing and applied to every service at once, in a space built from the work components the Social Security Act already names. In this new system, no service is an anchor, and no interested party's estimate of its own value enters the rules. Relativity becomes a property the schedule as a whole can be tested for, rather than an aspiration no process enforces.

The reason to state rules rather than strengthen the RUC is capacity. No committee, however expert, can hold 6,671 services in mind at once and check that each sits correctly against all the others. The current system therefore substitutes what is performable: valuing one service at a time, against a handful of comparators someone selects. Every local check used in the current system is a reasonable response to an impossible assignment.

Background

In 1987 Craig Reynolds demonstrated that the flocking of birds, a complex and coherent global pattern, requires no leader, no plan, and no bird that can see the whole flock.[17] It requires only that each bird follow three simple rules with respect to its immediate neighbors: keep clear of them, match their speed and heading, and stay with them.

---

[16] Proposed Rule, 91 Fed. Reg. at 43,953 (section II.H, Question 4).

[17] Reynolds CW. Flocks, Herds, and Schools: A Distributed Behavioral Model. *Computer Graphics*. 1987;21(4):25–34 (SIGGRAPH '87). Reynolds's terms for the three rules are collision avoidance, velocity matching, and flock centering.

Reynolds called his simulated birds "boids," and his three rules travel under that name. Coherence at the level of the flock is not designed. It emerges from every bird obeying local rules: separation—steer to avoid crowding your neighbors; alignment—steer toward your neighbors' average heading; and cohesion—steer toward your neighbors' average position. Relativity is the coherent pattern the flock produces: a global property that no participant can see, set directly, or inspect in one place. It can be produced, and it can be tested.

This is not an untested idea borrowed from birds. Industry has run consequential operations on stated local rules for decades. A quarter century ago the Harvard Business Review documented businesses applying these methods to real logistics, including cargo routing at Southwest Airlines.[18] Amazon's fulfillment network now operates more than one million robots coordinating in shared space.[19] The internet itself routes traffic this way: no entity sees the whole network, and global coherence emerges from every node obeying published local protocols. Emergent coordination from local rules is standard engineering in logistics, robotics, and communications. What this letter proposes is its application to a schedule of 6,671 relative values—a smaller flock than the ones industry already trusts to these principles.

The parallel to the RBRVS is exact in the respect that matters here. The schedule is the flock: each service is a bird, and the local rules it must satisfy can be defined. The work to develop the New RBRVS from the Boids Rules has three stages. Building the model gives every service a measured position in the "work-character space," defined below. It also requires a clear set of local rules like the ones Reynolds employed. Once all the codes are placed and the rules are set, running the model will let every value move toward consistency with its neighbors, all services at once. A third step, testing the schedule, measures how far each current value sits from consistency with the services near it, without moving anything.

Building the Model

Measuring the Work-Character Space

Every service must have a position assigned by the character of its work, so that "the services nearest this one" is a measured fact rather than someone's selection. I call the set of those positions the "work-character space." Today, "which services are most like this one" is answered when a survey-taker picks the comparators. Replacing that selection with measurement is the New RBRVS's critical first step.

---

[18] Eric Bonabeau and Christopher Meyer, "Swarm Intelligence: A Whole New Way to Think About Business," Harvard Business Review, May 2001, https://hbr.org/2001/05/swarm-intelligence-a-whole-new-way-to-think-about-business.

[19] Amazon, "Amazon launches a new AI foundation model to power its robotic fleet and deploys its 1 millionth robot," July 1, 2025, https://www.aboutamazon.com/news/operations/amazon-million-robots-ai-foundation-model.

The statute names two quantities. Time is already measured; the task is to give intensity a form that permits comparison. Intensity has three parts—mental effort and judgment, technical skill and physical effort, and psychological stress—and two services can carry identical time while differing sharply in the kind of demand they impose. Each service therefore needs coordinates, so that "nearest this one" is a distance in a multi-dimensional space rather than the difference between two one-dimensional RVUs.

The measurement already exists: the component ratings in the RUC's own survey records. Three decades of ratings sit in the RUC's record—the reports reached here span 2010 and 2014 through 2026—with two properties that dictate their handling. They are relative: every rating compares to a reference service, never an absolute score. They are also correlated and noisy: a service judged demanding on one component tends to be judged demanding on several, and any single rating carries substantial error.

To establish these measurements, I discarded the proposed values entirely and used only the component ratings, the relative, ordinal descriptions of the work. I ran principal component analysis on those ratings alone: the mental effort and judgment, technical skill, physical effort, and psychological stress comparisons. Neither time nor any payment figure entered the analysis. Given ratings that are relative, correlated, and noisy, the task is to turn many correlated measurements into a few stable coordinates. Principal component analysis is the standard tool for that task. When several measurements move together, it finds the smaller number of directions along which they vary, keeping the structure and discarding the redundancy. It fits nothing to payment and asserts nothing about value.

I found that two directions account for most of the variation in those ratings. The first is the movement the components share: the ratings tend to rise and fall together, so this direction captures whether a service was rated more demanding than its reference across the board. This becomes the first axis of the work-character space, and it reads as overall intensity. The second direction is the contrast the components do not share: some services rate high on mental effort and judgment relative to technical skill and physical effort; others the reverse. That contrast is the second axis—interpretive, cognitive demand at one end, technical, physical demand at the other. Anticoagulation management and electroencephalography fall at the cognitive end; bone marrow biopsy and fluoroscopic guidance at the physical end.

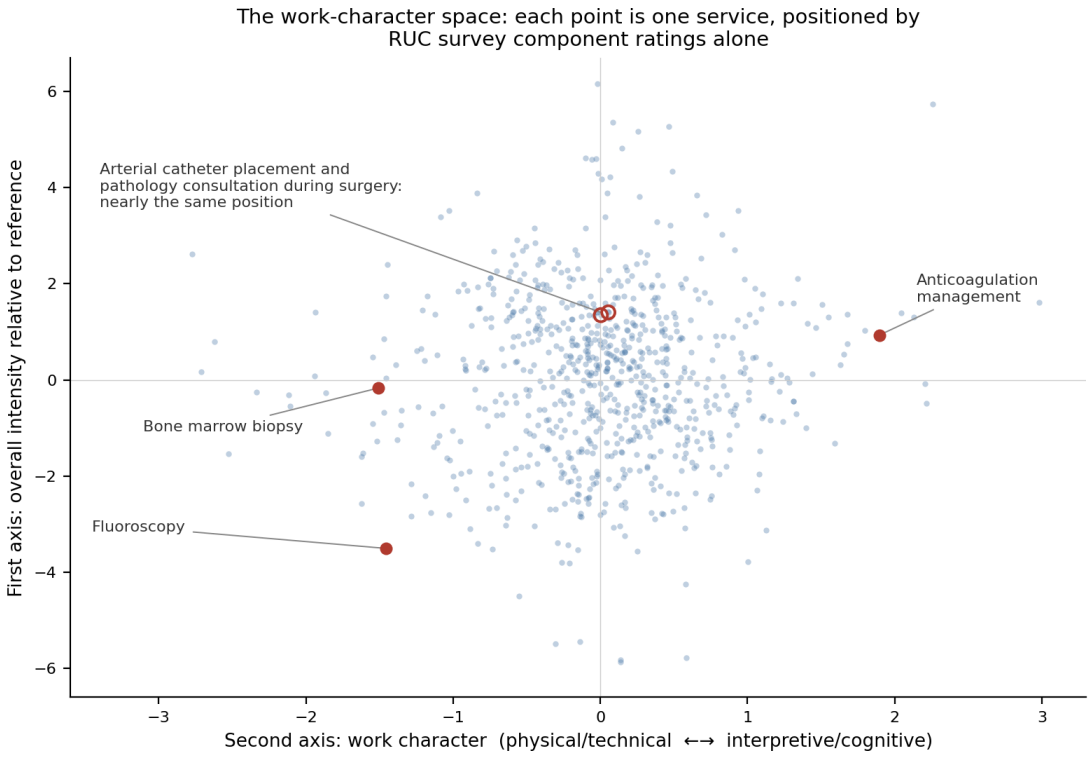


Figure 2: The work-character space

Figure 2. Each point is one of the 781 services whose survey component ratings are published—positions need only the component ratings, so this set is larger than the 282 with numeric intensity values—positioned by those ratings alone—no time and no payment figure enters the placement. The horizontal axis is the work-character contrast (physical/technical work to the left, interpretive/cognitive work to the right); the vertical axis is overall intensity relative to each service's own reference. Labeled points are discussed in the text. One extreme point falls outside the frame.

Nothing in my method specified the second axis. I never told the analysis that cognitive and physical work differ; it recovered the distinction from the data. Hsiao names mental effort and judgment, technical skill and physical effort, and psychological stress; and an analysis run without reference to that list recovers substantially the same structure. The dimensions of the work-character space are the system's own, confirmed rather than assumed. One caveat travels with the vertical axis: because every rating is relative to the service's own reference, it measures intensity against that reference rather than on an absolute scale.

One pair of services shows what the coordinates surface, and it is visible in Figure 2. An arterial catheter placement of about 30 minutes, billed most often by nephrologists, and a pathology consultation during surgery, about 25 minutes, land in nearly the same position in this space. On their face the two have little in common: one threads a catheter into an artery, the other reads tissue while the surgeon waits. The coordinates do not claim the services look alike. They record that the physicians who described each service, in

separate surveys, rated the work as nearly the same on both axes: similarly intense relative to its reference, and similarly balanced between interpretive and technical demand. Both are short services dense with judgment, performed with consequences attached. If those descriptions are right, the two are comparable work under the statute's own definition, and the gap in their values, 4.07 against 1.17, is a question the schedule should have to answer. If the descriptions are wrong, the pair is exactly what the clinical validation described below is built to catch. Either way, the comparison becomes a question that can be asked and answered. Nothing in the current process ever puts those two services side by side. A consistency rule would.

Time does not enter the space. It is the multiplier, as the statute defines it: each service keeps its own measured time, and similarity is judged on the character of the work alone.

The Rules

Given the work-character space, Reynolds's original triad—separation, alignment, and cohesion—maps onto the RBRVS with one structural difference. A Boids flock never lands. A fee schedule is meant to settle each year after policies are implemented. Two of the Boids Rules therefore translate into conditions a settled schedule either satisfies or violates, and can be tested today. The third governs only motion, and it belongs to the running stage.

Cohesion translates to local consistency: similar work carries similar intensity per minute—the work value divided by the intraservice time. A bird steers toward the center of the flockmates it perceives; a service's work value must be consistent with the intensity per minute of the services near it in the work-character space, applied to the service's own time. Time is the multiplier, as the statute defines it, so duration never masquerades as similarity: the services near this one are near in the character of the work, and each keeps its own time. This is the coherence condition. I built a consistency test on this principle and ran it against the existing fee schedule—a test, not a correction. Its findings are reported under "Testing the Fee Schedule," below.

Separation translates to rank order: within a family of related services whose own description attests an order—one service more extensive than another—the work values must keep that order. Birds may not pass through each other; values may not cross. An inversion is the collision. The rule anchors nothing, because an order is a claim about the description of the work, not about any value's level: holding that a repair with graft exceeds a repair without one fixes neither number, so no service becomes a reference point by participating in an order. The order claims exist today, in the valuation record's own language, service by service. Rank order is a candidate rule rather than a conclusion: extracting those attested orders at scale is unfinished work, and the rule stays subordinate to measured positions, because an order enforced between wrongly placed services would propagate their error.

Alignment has no analog at rest. Velocity matching is a rule about motion; in a settled schedule nothing moves, so no schedule can satisfy or violate it. Alignment belongs to the running stage: when values are adjusted, neighboring corrections move at matched rates, so a region of the schedule moves together rather than one service at a time. That is the orderly version of the updating that today happens code by code with no requirement that the results cohere.

The rules share one construction principle: a service's position in the work-character space comes from descriptions of the work itself—what the service is like to perform—and never from the value assigned to it. The assigned value is the thing being tested. If the test used it to choose the comparison set, the test could only agree with itself. There is no anchor, and there cannot be one. In flocking there is no lead bird. No service is privileged, and no declared reference code enters any of the rules as a fixed point. Anchoring to a small set of reference values is what allows one error to propagate through a family.

In defining the rule set, I found that most candidate rules were narrower statements of the two Boids Rules. For example, a family's internal coherence, an add-on staying beneath its base, and cross-specialty equivalence were absorbed into the rules above. Every rule I rejected fell to one test: it would have used the schedule's own values as an input, by preserving historical ratios, adopting a declared crosswalk as an equality, or fitting expected values to the current curve. The current schedule's relationships are the thing under test; they cannot also be an input.

Furthermore, nothing the RUC knows is discarded. Each kind of knowledge enters through a channel matched to what it can support: orders constrain, comparisons point, ratings position, clinicians calibrate. Only the rules move a value, and they move it in the open, where every stakeholder can see.

What is not Settled in the Method

The number of dimensions is the first open design question, and it is a working choice rather than a result. The working structure of this proposal uses two: overall intensity and the cognitive-physical axis, with time entering as each service's own multiplier rather than as a dimension. However, the question of how many axes are needed is empirical. Too few dimensions and genuinely different work is treated as similar; too many and the space becomes sparse, neighbors become distant, and local consistency loses its grip. The independent clinical validation described under "Testing the Fee Schedule," below, is designed to help settle it by probing for a dimension the current structure may be missing.

The weighting of the axes is also unsettled. Distance depends on how much each axis counts, so weighting determines which services qualify as neighbors and therefore what counts as an inconsistency. Equal weighting—a step of a given size along any axis counting the same—is a starting point rather than a finding. I state no weights here: weighting parameterizes distances in a space whose population is still changing.

Whether time belongs in the similarity judgment was the most consequential design question, and the statutory definition answered it. If work is time multiplied by intensity, time should not also be a coordinate defining similarity. The space is therefore built on work character alone; the rule tests intensity per minute; and each service keeps its own time as the multiplier. The trade-off is accepted deliberately: services of very different length will show up as comparable in this space, because the question the rule asks is not whether their total values match but whether their work carries a consistent intensity for the time it takes. I have retained a prior formulation with time as a third coordinate and raw work values tested as a documented sensitivity check.

The design also isolates a question the current process can only negotiate: whether intensity per minute is constant across duration—whether the demanding hours late in a long operation carry the intensity of the first. Because the rule separates the character of the work from its length, that question becomes measurable, and its answer belongs to evidence rather than advocacy.

Running the Model: How the New RBRVS would operate

In the New RBRVS, every service feels a pull toward the value the services near it imply, proportional to the gap between its current value and that implied value. All services move at once; the model iterates until local consistency holds everywhere it can hold; and no correction is final until every other service has finished moving. That simultaneity is what makes the result a schedule rather than a list of adjustments. This is the flock in motion, and the pattern it settles into is relativity.

Utilization behaves as mass in the model—in the flock, the heaviest birds turn slowest: a correction to a service performed ten million times a year moves more real spending than the same correction to a rare service, so high-volume services move more slowly. That is inertia, not influence: utilization changes what it costs to move a value, never a value's claim to be right, and it gives no service any pull on another.

There is no budget neutrality in the model I built. Today budget neutrality constrains every individual revaluation, which is what makes each one zero-sum and therefore contested. But budget neutrality plays no part in the correction itself. It enters afterward, as a funding policy that may be applied in full, in part, or not at all. Therefore budget neutrality may be applied here as a policy that adjusts payment after the model settles, rather than as a component of relativity.

This proposal splits two questions this program has always had to settle in one conversation: whether a service is valued correctly relative to other work, and whether Medicare is paying enough overall. The first is a coherence question with a testable answer. The second is a policy judgment belonging to CMS and to Congress. Conflating them is why every revaluation becomes a fight about adequacy, and every adequacy debate becomes a fight about relative values.

A schedule that corrects itself by rule gives CMS new capabilities. Suppose an Administration sets a priority—strengthening primary care in rural areas. Today pursuing it means hand-adjusting individual codes and defending each adjustment separately. With a running model, CMS could raise the values of the relevant services, offer a stipend, or adjust another lever, and test each option against the whole schedule before adopting any of them.

Every value becomes correctable, not merely contestable. Today an inconsistent value persists until someone brings it to a committee and wins the argument. In the New RBRVS, computation identifies a violating value and, once the corrective machinery is built and validated, corrects it in view of CMS, the specialty societies, and the public. Argument then attaches to the model, its parameters, and the policies applied to it, instead of to each of 6,671 values one at a time.

The running stage is a design. It has not been built or run, and this letter makes no claim to produce values for adoption.

Testing the Fee Schedule

The New RBRVS's local consistency rule is a condition the current schedule either satisfies or violates, so it can be tested today, before anything else is built. I built that test and ran it against the current physician fee schedule. Its object is the current schedule, not the proposal: the test holds every work value still and asks, service by service, whether the value is consistent with the intensity per minute of similar work—the standard the schedule itself claims to embody. The test works in three steps. For each service, it finds the services near it in the work-character space—every service within a stated distance. Those services form the "neighborhood": a recorded, auditable list, so that anyone can see exactly which services a value was checked against. The test then takes the intensity per minute those comparables carry, applies the service's own time, and compares the value that implies with the value assigned. The neighborhood belongs to testing only: when the model runs there is no list; influence fades continuously with distance, as a bird responds to whatever is near it. Discreteness serves verification; continuity serves motion.

Run against any schedule—this year's or any proposed revision—the test returns one of four statements for every service tested: 1) consistent with its comparables; 2) flagged, with the direction and size of the departure; 3) set aside for data quality, with the reason stated; or 4) without basis for local comparison, which is itself a finding about the schedule's evidence. It is a standing test, not a study. That is what it means for relativity to be checked rather than asserted—continuously, reproducibly, and in public.

I built everything that exists today from public records, and every piece can be traced to its source. Appendix B lists the sources; Appendix C summarizes the analyses and their results. Fee schedule values, times, and utilization come from CMS's published relative value files, work-time files, and Medicare utilization data for 2015 through 2026, and from

the proposed and final rules. The RUC material comes from the committee's own published meeting reports—thirty-nine in all—and from minutes posted publicly by the American Medical Association (AMA). The survey-form exhibits come from public rulemaking dockets at regulations.gov.

This test was run on the 781 services whose component ratings are published—12 percent of the schedule, and not a representative slice of it: they are shorter, lower-valued, and far less often 90-day surgical than the schedule as a whole. The direction is what survives that limitation, and it holds whether neighborhoods are defined with time, without time, or on time alone, which places the finding in the character of the work rather than in duration. Magnitude is not reported for the same reason.

The test found that the inconsistencies it detected are not randomly distributed. Surgical services tended to carry work values above those of the services nearest them in the work-character space. Cognitive services tended to carry work values below their comparators. A sensitivity run asked whether the findings survive different design choices—they do, in direction, across every setting tested.

There are limitations, as there are in finalizing the New RBRVS proposal, and they depend on data access and clinical input. As noted above, the work-character space requires independent clinical validation. If an entire region of the survey record is biased in one direction, the neighborhoods built from it are biased with it. Using only the ordinal ratings narrows the exposure, and cross-specialty neighborhoods keep any one specialty from positioning itself out of a finding, but neither closes it. The solution is direct: put the neighborhoods in front of practicing clinicians outside the valuation process and ask whether the services the model treats as comparable are, in their judgment, comparable work. Until that is done, the model tests. It does not generate values.

Reach is a second constraint. I built the reference network from the data that could be accessed—the recommendation records in the RUC's published meeting reports. It reaches 1,818 of the 6,671 services in the CY 2026 schedule, or 27%. The comparisons that would extend it exist in the RUC's own recommendation narratives, where the committee's justifications routinely name comparator codes beyond the declared references. Extending its reach requires processing the recommendation narratives for the remaining services.

The parameters are working choices. Neighborhood size and the tolerance that defines an inconsistency are starting points, not derived quantities. The direction of the findings is stable across those choices, and no specific violation rate is. That is why this letter reports none.

What Next

Three steps remain to build the New RBRVS, and they run through clinicians.

- First, a calibration survey: clinicians outside the valuation process judge, without payment values attached, whether the services the model treats as comparable are comparable work. The survey validates the work-character space, probes for a dimension the ratings may have missed, and places the services the published record cannot reach. Appendix D reproduces a sample item from the instrument, which is built and needs to be fielded.
- Second, scale. The committee's order claims—the attested orders the rank order rule runs on—are extracted from its own published narratives, extending the tested territory toward the services the reference network never reaches.
- Third, and only after validation, the running stage. The order is deliberate: the model earns each stage before it claims the next. The record is public and the sequence is stated.

Figure 3 shows the full sequence and the build state of each stage.

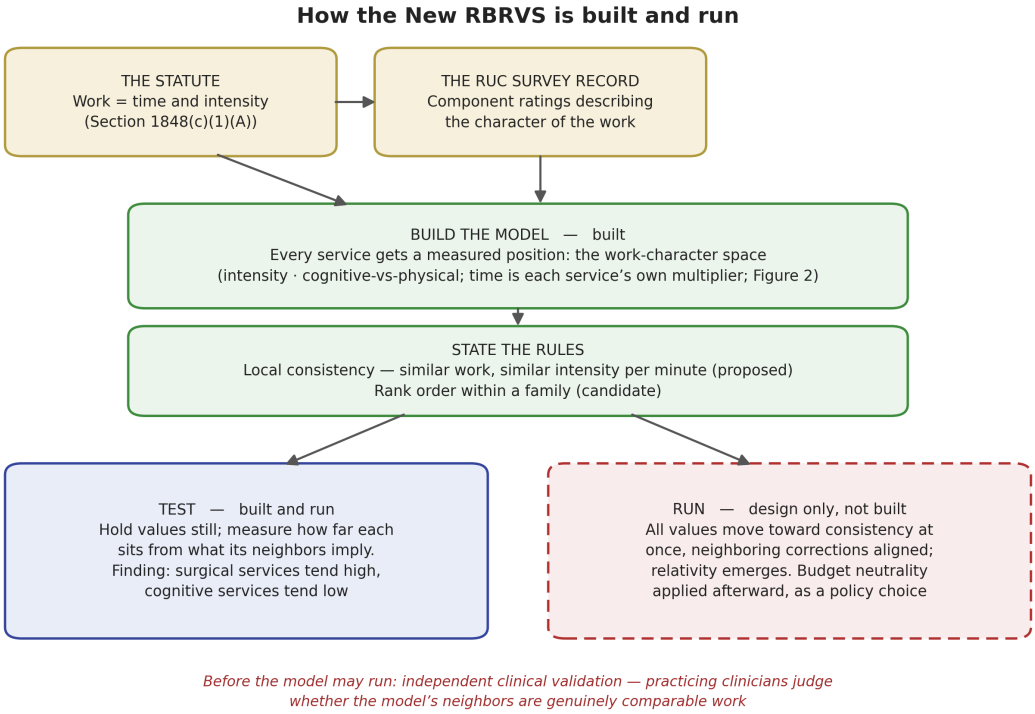


Figure 3: How the New RBRVS is built and run. Solid stages are built; the running stage is design only and gated behind clinical validation.

Adoption and Innovation

A stated consistency rule makes valuing a genuinely new service a test against comparable work rather than a negotiation. That shortens the path for services with no natural specialty champion and no established constituency, because a defensible value no longer depends on who is available to advocate for it. The same rule imposes a discipline in the other direction: a new service whose proposed value cannot be justified against comparable

work would no longer obtain that value through advocacy alone. Both effects serve innovation, because the first rewards genuinely new work on its merits and the second stops rewarding representation as if it were merit.

The same rules give a new service a stated on-ramp. A new service needs two things for a defensible value on day one: a position, assigned by the character of its work, which clinicians can judge from the service itself; and its own measured time. Its value follows from the intensity per minute of the services nearest it, applied to its own time. No constituency is required, and no negotiation. Because time is the multiplier, the value would also track measured time as technique matures. The learning-curve correction that today requires a contested revaluation would happen by arithmetic. That is one reason this letter supports the empirical time data proposal at section II.D: the same data that validates existing times would keep new services' values current.

One safeguard belongs in the design from the start. A new service should be judged by the schedule before it shapes the schedule. Until its evidence matures, a new service would carry a value derived from its neighbors without entering any other service's comparison set. Emerging technology receives a defensible value immediately; the uncertainty that accompanies it does not propagate. I state this as an expectation of the design rather than a demonstrated result. The mechanism requires testing once the model runs, and that testing is among the work I would welcome the opportunity to show CMS.

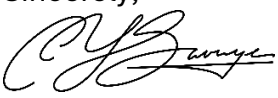
Closing

These comments rest on evidence about work relative value units only. Nothing in my findings endorses or indicts current practice expense values. No service is described here as overpaid or underpaid. These findings concern internal consistency among relative values. Payment adequacy is a distinct question governed by the conversion factor—the dollar multiplier that turns relative value units into payment—and budget neutrality, and it is not measured here. Services are described as sitting high or low relative to comparable services, and deliberately not as overvalued or undervalued, because those words assert an absolute judgment this work does not make.

I will provide the code and the analysis-ready data to CMS staff on request, and I would welcome the opportunity to discuss the methodology, the data handling, and the findings in whatever form is useful to CMS's consideration of these questions.

Thank you for considering these comments. Should you have any questions, please contact me at (202) 680-8985 or csavage@shpconsulting.llc.

Sincerely,



Courtney Yohe Savage, MPP
Principal
Savage Health Policy, LLC

CPT Descriptor

RELATIONSHIP OF CODE BEING REVIEWED TO TOP TWO KEY REFERENCE SERVICES:
Compare the pre-, intra-, and post-service time (by the median) and the intensity factors (by the mean) of the service you are rating to the top two chosen key reference services listed above. **Make certain that you are including existing time data (RUC if available, Harvard if no RUC time available) for the reference code listed below.**

Number of respondents who choose Top Key Reference Code: 25 **% of respondents:** 21.9 **%**

Number of respondents who choose 2nd Key Reference Code: 23 **% of respondents:** 20.1 **%**

TIME ESTIMATES (Median)

	CPT Code: 36901	Top Key Reference CPT Code: 36200	2nd Key Reference CPT Code: **36251**
Median Pre-Service Time	26.00	41.00	38.00
Median Intra-Service Time	25.00	30.00	45.00
Median Immediate Post-service Time	15.00	20.00	30.00
Median Critical Care Time	0.0	0.00	0.00
Median Other Hospital Visit Time	0.0	0.00	0.00
Median Discharge Day Management Time	0.0	0.00	0.00
Median Office Visit Time	0.0	0.00	0.00
Prolonged Services Time	0.0	0.00	0.00
Median Subsequent Observation Care Time	0.0	0.00	0.00
Median Total Time	**66.00**	**91.00**	**113.00**
Other time if appropriate			

INTENSITY/COMPLEXITY MEASURES
(of those that selected Key Reference codes)

Survey respondents are rating the survey code relative to the key reference code.

Intensity & Complexity Rating Scale: (much less= -2.00, somewhat less= -1.00, identical= 0.00, somewhat more= 1.00, much more= 2.00)

	Top Key Ref Code	2nd Key Ref Code
Mental Effort and Judgment (Mean)		
The number of possible diagnosis and/or the number of management options that must be considered	-0.20	0.39
The amount and/or complexity of medical records, diagnostic tests, and/or other information that must be reviewed and analyzed	0.12	0.39
Urgency of medical decision making	0.00	0.17

Technical Skill/Physical Effort (Mean)

Technical skill required	-0.16	-0.13
Physical effort required	-0.12	0.17

Psychological Stress (Mean)

The risk of significant complications, morbidity and/or mortality	-0.48	-0.09
Outcome depends on the skill and judgment of physician	0.00	0.22
Estimated risk of malpractice suit with poor outcome	-0.20	-0.13

INTENSITY/COMPLEXITY MEASURES	Top Key Ref Code	2nd Key Ref Code
Time Segment (Mean)		
Overall intensity/complexity	-0.20	0.35

Additional Rationale and Comments

Describe the process by which your specialty society reached your final recommendation. *If your society has used an IWPUT analysis, please refer to the Instructions for Specialty Societies Developing Work Relative Value Recommendations for the appropriate formula and format.*

The additional rationale below is the original rationale submitted by the specialty society(ies) prior to the RUC meeting and does not necessarily represent the rationale for the RUC recommendation. To view the RUC's rationale, please review the separate RUC recommendation document.

Background

CPT code 36870 was identified by CMS as potentially misvalued on the high expenditure screen. It was sent to CPT to create a new bundled family to describe intervention and maintenance of hemodialysis access. The specialty societies (SVS, SIR, ACR, and RPA) presented the new family of codes at the October 2015 CPT meeting. The new family consists of nine new codes to describe a fistulogram (36901), fistulogram with angioplasty (36902), fistulogram with angioplasty and stent (36903), thrombectomy (36904), thrombectomy with angioplasty (36905), and thrombectomy with angioplasty and stent (36906). There are also three add on codes to describe central venous angioplasty done through the hemodialysis access (+36907), central venous stent through the hemodialysis access (+36908), and coil embolization of the hemodialysis access (+36909).

Methodology

A multi-disciplinary expert panel including SVS, SIR, ACR, ASIN, and RPA was convened. A survey was distributed randomly to the members of the representative societies. There were 114 surveys completed. The surveys were reviewed by the multi-specialty group and determined to be valid and reflective of the work and intensity involved.

Work RVU Recommendation for 36901

We are recommending the 25th percentile survey value of 3.72 RVW for 36901 which is the existing value.

Pre- and Post-Service Time Packages

The new CPT code 36901 will replace the existing 36147 code for fistula access and imaging of the fistula and outflow. Medicare utilization data from 2014 shows that this procedure is performed in the physician office

RELATIONSHIP OF CODE BEING REVIEWED TO TOP TWO KEY REFERENCE SERVICES:
Compare the pre-, intra-, and post-service time (by the median) and the intensity factors (by percent distribution) of the service you are rating to the top two chosen key reference services listed above. **Make certain that you are including existing time data (RUC if available, Harvard if no RUC time available) for the reference code listed below.**

Number of respondents who choose Top Key Reference Code: 33 **% of respondents:** 34.0 **%**

Number of respondents who choose 2nd Key Reference Code: 14 **% of respondents:** 14.4 **%**

TIME ESTIMATES (Median)

	CPT Code: 33016	**Top Key Reference CPT Code:** 32557	**2nd Key Reference CPT Code:** **37236**
Median Pre-Service Time	46.00	22.00	31.00
Median Intra-Service Time	30.00	30.00	90.00
Median Immediate Post-service Time	20.00	15.00	30.00
Median Critical Care Time	0.0	0.00	0.00
Median Other Hospital Visit Time	0.0	0.00	0.00
Median Discharge Day Management Time	0.0	0.00	0.00
Median Office Visit Time	0.0	0.00	0.00
Prolonged Services Time	0.0	0.00	0.00
Median Subsequent Observation Care Time	0.0	0.00	0.00
Median Total Time	**96.00**	**67.00**	**151.00**
Other time if appropriate			

INTENSITY/COMPLEXITY MEASURES
(of those that selected Key Reference codes)

Survey respondents are rating the survey code relative to the key reference code.

Top Key Reference Code	**Much Less**	**Somewhat Less**	**Identical**	**Somewhat More**	**Much More**
Overall intensity/complexity	0%	0%	6%	61%	33%

Mental Effort and Judgment	**Less**	**Identical**	**More**
• The number of possible diagnosis and/or the number of management options that must be considered • The amount and/or complexity of medical records, diagnostic tests, and/or other information that must be reviewed and analyzed • Urgency of medical decision making	0%	15%	85%

Technical Skill/Physical Effort	Less	Identical	More
Technical skill required	0%	6%	94%
Physical effort required	0%	45%	55%

Psychological Stress	Less	Identical	More
• The risk of significant complications, morbidity and/or mortality • Outcome depends on the skill and judgment of physician • Estimated risk of malpractice suit with poor outcome	0%	6%	94%

2nd Key Reference Code	Much Less	Somewhat Less	Identical	Somewhat More	Much More
Overall intensity/complexity	0%	0%	21%	50%	29%

Mental Effort and Judgment	Less	Identical	More
• The number of possible diagnosis and/or the number of management options that must be considered • The amount and/or complexity of medical records, diagnostic tests, and/or other information that must be reviewed and analyzed • Urgency of medical decision making	7%	36%	57%

Technical Skill/Physical Effort	Less	Identical	More
Technical skill required	0%	29%	71%
Physical effort required	0%	21%	79%

Psychological Stress	Less	Identical	More
• The risk of significant complications, morbidity and/or mortality • Outcome depends on the skill and judgment of physician • Estimated risk of malpractice suit with poor outcome	0%	14%	86%

Additional Rationale and Comments

Describe the process by which your specialty society reached your final recommendation. *If your society has used an IWPUT analysis, please refer to the Instructions for Specialty Societies Developing Work Relative Value Recommendations for the appropriate formula and format.*

The additional rationale below is the original rationale submitted by the specialty society(ies) prior to the RUC meeting and does not necessarily represent the rationale for the RUC recommendation. To view the RUC's rationale, please review the separate RUC recommendation document.

Appendix B — Data sources and access

Source	What it provided	Where obtained
CMS Physician Fee Schedule relative value files, CY 2015–2026	Work RVUs, global periods, status indicators for every service	cms.gov, public downloads
CMS physician work-time files	Pre-, intra-, and post-service time assumptions	cms.gov, public downloads
Medicare utilization data, 2015–2026	Service-level volumes	CMS public use files
CY 2027 proposed rule and prior PFS rules	The policy statements quoted throughout	federalregister.gov; regulations.gov
RUC meeting reports — thirty-nine in all: thirty-six annual reports, October 2014 through May 2026, and three Five-Year Reviews	Valuation narratives, declared reference services, survey component ratings, stated justifications	AMA publications, publicly posted
RUC meeting minutes	Time and Intensity Workgroup and Research Subcommittee records of the 2016–2017 survey-instrument change	AMA website, publicly posted
Survey-form exhibits	The RUC survey instrument before and after the 2017 change	Comment records on public rulemaking dockets, regulations.gov
Social Security Act § 1848	The statutory definition of physician work	42 U.S.C. § 1395w-4

The RUC record is published as narrative PDFs, and purpose-built software recovered its structure, storing alongside every extracted claim the exact passage it came from. Where an extraction's reliability is reported, the number comes from re-reading source documents against the output — never from the software grading itself.

Appendix C — Analyses conducted

The question	What was done	What it showed
How much of a service's work value can be predicted from its time alone?	For the 6,415 services in the CY 2026 schedule that carry a recorded intraservice time, each service's work RVU was compared with it, and the strength of the relationship was measured.	Time alone explains 81% of the variation in work values—86% in logarithms—and the strength of the relationship has not moved since 2015.
Does the schedule pay a consistent amount of work value per minute?	Each service's work RVU was divided by its intraservice time, and the spread of that rate was measured across the schedule (90th percentile against 10th).	The rate varies by a factor of 3.7. That variation is where intensity should be visible.
Does measured intensity add anything to what time already predicts?	For the 282 services whose survey intensity ratings are published, values were predicted from time alone, and then from time plus the ratings, and the two predictions were compared.	Adding measured intensity improves the prediction by about one tenth of one percentage point (log 0.855 to 0.856; levels 0.922 to 0.923; n = 282).
Can intensity be measured from the survey record at all?	The survey's component ratings for 781 services — and nothing else: no times, no payment values — were analyzed with a standard statistical method (principal component analysis) that finds the directions in which ratings move together.	Two directions: overall intensity, and a cognitive-versus-physical contrast that nothing in the method asked for.
Are work values consistent among services whose work is similar?	Each service's value was compared with the values of the services nearest it, and the test was re-run three ways: with time helping define nearness, without it, and with nearness defined by time alone.	The direction never changes: surgical services tend to sit above their comparables, cognitive services below. No violation rate is quoted, deliberately — the rate moves with the settings; the direction does not.
How much of the schedule does the committee's own	Every declared key reference service was extracted from	1,818 of 6,671 services — 27%. The other 73% is not reached by a declared key

comparison record reach?	the published recommendation records.	reference in the meeting reports covered here.
Do the committee's justifications stay within the survey's declared references?	356 valuation recommendations across five reports (2010–2025) were checked by hand against the source documents.	90% relied on comparator services beyond the declared reference. Agreement between the hand check and the coded output was 18 of 20, and the figure is reported with that caveat.
What happened to the survey's intensity questions in 2017?	The committee's own minutes and the survey forms filed on public dockets were read.	Eight component questions were collapsed into fewer. The stated reasons were administrative — reviewer burden and response rate — and the validity concern the committee recorded in the same passage is not addressed in the minutes of the change.

Appendix D — Sample calibration item: Clinical Intuition Survey

The calibration survey places the model's comparability judgments in front of practicing clinicians outside the valuation process. The instrument was built to the OMB Standards and Guidelines for Statistical Surveys and the survey-methodology literature. Respondents are never shown payment values, and no question asserts the model's conclusion before asking for judgment. The item below is Comparison 5 of 12, reproduced verbatim from the current instrument, with build annotations omitted; service descriptions are CMS's consumer-friendly descriptors. The instrument has not yet been fielded; a pretest precedes fielding.

Below are two procedures. They have similar intraservice times (30 and 30 minutes). Please judge for yourself whether the physician work they require is similar or different.

	Procedure A	Procedure B
CPT Code	28820	99283
Description	Amputation of toe at joint between forefoot and toes	Emergency department visit with low level of medical decision making
Typical specialty	Podiatry	Emergency Medicine
Intraservice time	30 minutes	30 minutes

Q0. How familiar are you with these two procedures?

- ☐ I perform or interpret both of them
- ☐ I perform or interpret Procedure A but not Procedure B
- ☐ I perform or interpret Procedure B but not Procedure A
- ☐ I perform or interpret neither, but I am familiar with both from practice
- ☐ I am not familiar with one or both of them

Q1a. Compared with one another, how demanding is the physician work in these two procedures?

- ☐ Procedure A is much more demanding
- ☐ Procedure A is somewhat more demanding
- ☐ About equally demanding
- ☐ Procedure B is somewhat more demanding
- ☐ Procedure B is much more demanding
- ☐ I do not know enough about one or both of these procedures to judge

Q1b. How similar is the kind of work the two procedures require — the mix of interpretive and diagnostic demands and physical and technical demands?

- ☐ Very similar
- ☐ Somewhat similar
- ☐ Somewhat different
- ☐ Very different

- ☐ I do not know enough about one or both of these procedures to judge

Q2. How confident are you in your two answers above about this pair?

- ☐ Very confident
- ☐ Fairly confident
- ☐ Not very confident
- ☐ Not at all confident

Q3. Which of these, if any, are meaningfully different between the two procedures? Select all that apply. Select none if none of them are.

- ☐ Consequences of error — how serious the outcome of a mistake would be
- ☐ Patient complexity — how sick or complicated the patients typically are
- ☐ Depth of specialization — how much advanced training the procedure demands within its own specialty
- ☐ Urgency — how often it must be done under time pressure
- ☐ Physical setting — where the procedure takes place
- ☐ Kind of work — interpretive and diagnostic demands versus physical and technical demands
- ☐ Duration — how long the procedure takes
- ☐ Other — please describe

Q4. Any additional observations about this comparison? (Optional)

Q1a and Q1b separate the two judgments the model makes — how demanding, and what kind of work — so agreement is measured on each construct rather than on a single yes/no. Q3, shown to every respondent with its order randomized, is the probe for a dimension of work the model's space may not carry. Payment values appear nowhere in the instrument.