

Online Appendix for Detecting Potential Overbilling in Medicare Reimbursement via Hours Worked: Comment

Brett Matsumoto[†]

Bureau of Labor Statistics[‡]

A Corrections to the Procedure Time Estimation Methodology

In this paragraph, I summarize the methodology to estimate procedure times as it is described in Fang and Gong (2017). They use 2 sets of timed codes to estimate the time needed for the set of codes for which no time measure is available. The first set of timed codes are codes where the AMA guidelines specify an amount of time. The second set of timed codes are from Zuckerman et al. (2014). Let T_0^{Tot} denote the total time for the timed codes and T_0^{Int} denote the intra-service time for the timed codes. The authors then generate estimates of intra-service and total time using a couple of different methods. The first method takes the time for the timed codes and divides by the work RVU to get a measure of time per work RVU for each category of code. Then, the work RVUs of the untimed codes are multiplied by the time per RVU to get an estimate of total time. Denote this estimate as T_1^{Tot} and T_1^{Int} . The second method regresses time on the wRVU for the timed codes (allowing for the coefficient on wRVU and the intercept to vary across code groups). The regression coefficients are then used to predict the times T_2^{Tot} and T_2^{Int} . The estimate for intra-service time (\hat{T}^{Int}) is the minimum of T_0^{Int} , T_1^{Int} , and T_2^{Int} . The estimate for total time is $\hat{T}^{Tot} = \max(\hat{T}^{Int}, \min(T_0^{Tot}, T_1^{Tot}, T_2^{Tot}))$. In other words, it is the minimum of the total time estimates unless the minimum of the total time estimates is less than the intra-service time estimate.

First, to clarify the discussion in the paper, for the first set of timed codes they have 2 different intra-service time values (typical time and expected time) and no total time value. The intra-service time is defined as the minimum of the typical and expected time (more on this below). So T_0^{Int} includes both sets of timed codes, which are used in the estimation of the intra-service time for the untimed codes. T_0^{Tot} *only* includes the timed codes from Zuckerman 2014. It is these times which are used to estimate the total time for the untimed codes. Since all of the analysis in the paper uses total time as the measure, the AMA times provide almost no information in the estimation of untimed codes. The only place they enter the estimate for total time is in the last step where total time is replaced with intra-service time if intra-service time is greater than the estimated total

[†]Contact: matsumoto.brett@bls.gov

[‡]The views and opinions expressed in this paper are my own and do not reflect the views of the BLS, Department of Labor, or the Federal Government.

time. This final step includes replacing the missing values for total time for the AMA codes with their intra-service time.

The authors make the following mistakes when implementing this procedure:

- For the AMA timed codes, the authors never take the minimum of the expected and typical times. They simply rename the expected time variable from the inputted spreadsheet as intra-service time. There are 10 codes for which the typical time is less than the expected time.
- The authors do not generate a predicted time for all procedures for the two estimation methods. When calculating T_1^{Int} and T_1^{Tot} , no estimate is calculated for the timed codes. When T_2^{Int} and T_2^{Tot} are calculated, the only estimates that are calculated are for the timed codes. The authors have clarified that they only intended to exclude the first set of timed codes when calculating the estimated time using each of the estimation procedures. So the authors' code doesn't calculate an estimate of T_1 for the Zuckerman timed codes, but should. For T_2 an estimate should be calculated for the untimed codes and the Zuckerman timed codes.
- Despite the intention to exclude the first type of timed codes, the authors' code generates a total time estimate for the regression based method for these codes.
- There are a couple of typos in the authors' spreadsheet for the timed codes from Zuckerman 2014. The code in Zuckerman 2014 is "44204", which is entered in the authors' spreadsheet as "44203". For code 22612 the authors have an intra-service time of 160, but the actual value is 150. These typos have minimal impact on the procedure time estimates and become irrelevant after correcting the next issue.
- An incorrect source of timed codes is used. The values in Zuckerman 2014 were the current AMA RUC times. The final results for the Urban Institute physician time project were reported in Zuckerman 2016.

The Table 1 summarizes the sources of the time estimates. The final total time estimate for each code can come from the time per RVU estimate (T_1^{Tot}), the regression estimate (T_2^{Tot}), the original time estimate for timed codes (T_0^{Tot}), the intra-service time estimate (\hat{T}^{Int}), or if there are no time estimates (all missing) the total time is set to zero. Since the authors' method did not generate time estimates for T_2 for untimed codes, the vast majority of time estimates were the values calculated for T_1 . Correcting the authors' methodology to take the minimum of multiple time estimates causes there to be more balance over the different estimates. Table 2 presents the summary statistics for the procedure times with and without the corrections. The average procedure time decreases by about 10 minutes among the 5393 procedure codes as the result of correcting the methodology. Using the final data from the Urban Institute study (Zuckerman et al., 2016) has less of an impact compared to correcting the methodology. The average procedure time falls by approximately 1 minute. The increase in zero time codes when using the new data is due to the regression based estimation method predicting more negative time values for radiology codes.

Table 1: Source of the Service Time Estimates

Source of \hat{T}^{Tot}	Original Paper	With Corrections	With Corrections and New Data
T_0^{Tot}	0	47	19
T_1^{Tot}	4,839	1,954	2,893
T_2^{Tot}	100	2,523	1,624
\hat{T}^{Int}	17	352	340
All missing	437	517	517

Notes: Source of the Total time estimate used in the calculation of physician working times. T_0^{Tot} includes the set of timed codes. T_1^{Tot} is the estimate using the group average method. T_2^{Tot} is the estimate using the regression method, and \hat{T}^{Int} is the intra-service time estimate.

Table 2: Summary of Service Times with Corrections

Time Estimates	Mean	SD	Median	Min	Max	# zero time codes
Original Method and Data	109.44	176.6	37.24	0	1850.3	1348
With Corrections	98.71	169.5	29.05	0	1826.2	1440
With Corrections and New Data	97.62	164.6	25.12	0	1767.8	1816

Notes: Total service time for the 5393 hcpcs codes included in the analysis.

B Service Count Adjustment Detailed Procedure

The physician fee schedule has a line for each hcpcs code representing the full service as well as a line for the modifiers for discontinued procedure (modifier code 53), professional component only (code 26), and the technical component only (TC) for services where those modifiers are applicable. Each of these code/modifier combinations has an assigned work RVU, which I use to estimate a procedure time based on the share of code’s total work RVU represented by the modifier (e.g., if the discontinued procedure modifier had half of the total code’s work RVU the time associated with the -53 modifier would be half the total time for the procedure).

Table 3 summarizes the other service count adjustments in the order in which they are performed. The first step is to set the service count to 1 for 10 and 90 day global surgical procedures that cannot have a service count greater than 1 by definition (unless it is as a bilateral procedure which is accounted for later).¹ There are 2 ways to indicate bilateral procedures. There is a bilateral procedure modifier code (50), but bilateral procedures can also be reported using both the left side (LT) and right side (RT) modifier codes. There are 3 cases for bilateral procedures. Some procedures are defined as bilateral procedures. These codes are not eligible for the bilateral adjustment because the adjustment is built into the RVU for the code. I set the service count to 1 for these codes. The other cases for bilateral procedures are a 150% and 200% adjustment. The 150% adjustment can be thought of as a multiple procedure adjustment for the second side. The 200% adjustment treats each side as being an entirely separate procedure. The bilateral adjustment can increase or decrease the original line service count depending on whether the physician bills the procedure with a quantity of 1 or a quantity of 2 (the claim is processed and paid the same regardless of which quantity is entered). For each code, the fee schedule has a breakdown

¹This adjustment is also performed on a handful of non-surgical procedures that have an assumed quantity of 1.

Table 3: Service Count Adjustments

Adjustment	Modifier Code	Fee Schedule Identifier	Service Count Adjustment
Surgical Procedure		Global days equal to 10 or 90	Set service count to 1
Bilateral Procedure	50 or LT & RT	Not eligible	Set service count to 1
		150% adjustment	Set service count to 1.5
		200% adjustment	Set service count to 2
Pre-Service Only	56	percent pre service	Set service count to pre-service share
Intra-Service Only	54	percent intra-service	Set service count to intra-service share
Post-Service Only	55	percent post-service	Set service count to post-service share
Multiple Procedure	51	Eligible for MP adjustment	Divide line service count by 2

of the percent of the total procedure corresponding to the pre-, intra-, and post-service periods. Physicians who only provide partial service are paid the corresponding percent of the total fee for the code. In these cases, the quantity is set to the proportion of the total service provided. Finally, claims submitted with the modifier code for multiple procedure (code 51) which are eligible for the adjustment have the line service counts divided by 2.

The goal of this procedure is to obtain a more accurate count of services for the purposes of calculating total physician time. These adjustments cover the most prominent modifier codes that indicate partial service, and are not exhaustive of the possible adjustments one could make. It is also possible that this procedure will fail to pick up non standard ways of indicating partial service.

C Using the 5% Beneficiary Sample to Estimate Total Working Time

Just looking at the 5% sample, there is strong evidence that the originally flagged physicians are disproportionately affected by the service count corrections. In order to quantify how much this disproportionate impact affects the propensity to be flagged requires estimating the time for 100% of claims using the time estimates in the 5% sample. In this section I show that the method used to generate an estimate of total time using the 5% sample and the unadjusted service counts generates a reasonable approximation to total time in the Utilization data. Additionally, I present simulation evidence that suggests that the weaknesses of using the 5% claims to estimate total hours worked make it MORE difficult to detect a decrease in flagging propensity as the result of using adjusted instead of the unadjusted service counts.

C.1 The Total Time Estimate versus Actual Time

Let T and A denote the total time and allowed charges in the full sample, and let t_5 and a_5 denote the total time and allowed charges in the 5% sample. The estimate of total time using the time in the 5% sample is:

$$T^* = t_5 * \frac{A}{a_5} \tag{1}$$

Table 4: Summary Statistics for Different Total Time Estimates

Data Source	Service Counts	Service Times	Mean Time	SD	Median	Min	Max
Utilization	Unadj	Original	10.01	15.68	5.268	0	948.5
Utilization	Unadj	New	8.449	13.75	4.240	0	857.6
Claims	Unadj	Original	11.49	17.83	6.348	0	1517.1
Claims	Unadj	New	9.744	15.60	5.158	0	1291.1
Claims	Adj	Original	10.75	14.25	6.116	0	713.9
Claims	Adj	New	9.174	12.66	4.983	0	780.8

Notes: Sample includes the 608,050 physicians who appear in both the Utilization and Claims data. Adjusted service counts attempt to correct for common sources of overcounting in the claims data. Original service times are the service time estimates used in Fang and Gong (2017). New service times correct for errors in the estimation methodology and source of time data.

Table 5: Correlation Matrix for Different Total Time Estimates

Data Source / Service Counts / Service Times	1	2	3	4	5	6
1. Util/Unadj/Old	1.000					
2. Util/Unadj/New	0.948	1.000				
3. Claims/Unadj/Old	0.913	0.867	1.000			
4. Claims/Unadj/New	0.867	0.915	0.955	1.000		
5. Claims/Adj/Old	0.865	0.820	0.843	0.804	1.000	
6. Claims/Adj/New	0.810	0.872	0.794	0.856	0.940	1.000

If the ratio of total time to charges is the same in the 5% sample and the 100% sample, then $T^* = T$. As an estimate of T , T^* is biased ($E[T^*] \neq T$ unless $Cov(t_5, \frac{1}{a_5}) = 0$) and is a noisy estimate of T . Any sampling variability in the ratio of time to allowed charges will be amplified by the transformation (since the procedure assumes the same ratio in the rest of the claims not included in the 5% sample). Since the Utilization data at the procedure level does not include 100% of claims, the total time calculated from that data will be less than the true T . Table 4 presents the summary statistics for the distributions of total time in the Utilization data and the estimates of T^* using both adjusted and unadjusted service counts for the physicians who appear in both. The results are also reported for both sets of time estimates. The estimates of T^* using unadjusted service counts in the claims data have a higher mean and median than the estimates from the Utilization data for both sets of service times. The distribution of T^* also has higher variance, which suggests a fatter right tail.

Table 5 presents the correlation matrix between the 6 different estimates of total time. The correlation between the estimated working time using the Utilization service level data and the unadjusted service counts in the 5% claims data is 0.913 using the original procedure times and 0.915 using the new procedure times.

C.2 Impact of Post-Service Care only Codes on the Claims Data Estimate

Table 6 looks more closely at the hours estimates from the Utilization data and the claims data with unadjusted service counts using the original service time estimates for both. The summary statistics

from these distributions are presented for subsamples of physicians as well as the correlation between the hours estimates. The top panel breaks the sample of all physicians into 2 groups based on whether they bill one of the 3 common post-service care only codes in the Utilization data (see Table 2 in the main text). The 3 common post-service care only codes are 66984, 66821, and 66982. These are codes that can be billed with very high service counts (up to 90) for a single patient. Since each patient has a 5% chance of appearing in the sample, physicians may not perform these procedures on enough patients for any to appear in the claims sample. The allowed charges for post-service care only are relatively low, so the estimated times from the claims data will be much lower than the estimated times in the Utilization data if none of the post service care codes is present in the claims data.

The doctors who perform the post-service care only codes have a much higher estimate of hours worked. The claims data estimates tend to be higher than the Utilization data estimate and have a higher variance. For doctors that do not bill any of the 3 common post-service care only codes, the correlation between the time estimates from the Utilization and claims data is 0.948. For the post-service care doctors, the correlation is only 0.800. The next panel compares the hours estimates for flagged and unflagged physicians who work over 20 hours per week in the Utilization data. The correlation between the Utilization and claims hours estimates is 0.813 for the unflagged and over 20 hour physicians and 0.721 for flagged physicians. However, breaking these groups up further based on whether or not the physicians billed any of the 3 common post-service care only codes (bottom panel) shows that the relatively low correlation between the hours estimates in the Utilization and claims data is due entirely to the doctors who perform at least one of these procedures in the Utilization data.

Table 7 shows the summary statistics aggregating across the 3 common post service care only codes among physicians in the Utilization data. The variables include the total service counts associated with these 3 codes, the number of unique beneficiaries, the total allowed charges (this equals the total paid to the physician by Medicare, patient, and 3rd party payers), the hours per week calculated from the total service counts in the Utilization data, and the implied wage (total allowed charges/total hours).

There are 15,154 physicians who bill at least one of these services in the Utilization data. The median physician has a total service count among these 3 procedures of 112 and the median number of beneficiaries is 58. Among the physicians flagged in either the Utilization data or the claims data (or both), the ratio of service counts to beneficiary is much higher. This indicates that these procedures are being billed with a service count greater than 1. It is possible for a single individual to have more than one of these 3 procedures from the same physician or a repeat procedure outside of the 90 day global billing window, but this is not a plausible explanation for the service count to beneficiary ratio observed among many of these physicians. Two of the 3 procedures are cataract surgeries that would not both be covered for the same individual and same eye. It is also physically impossible for cataract surgery to be performed more than once on an eye since the procedure

Table 6: Comparing Utilization and Claims Data Hours Estimates

Group	N	Data	Mean	SD	Median	Min	Max	Correlation
Non Post Care Doctors	592,903	Utilization	9.400	13.58	5.068	0	948.5	0.948
		Claims	10.87	14.93	6.151	0	955.5	
Post Care Doctors	15,147	Utilization	33.86	45.49	21.40	0.400	740.6	0.800
		Claims	35.48	58.67	20.64	0.258	1517.6	
Unflagged and > 20 hrs/week	80,972	Utilization	34.46	14.97	29.35	20.00	99.99	0.813
		Claims	37.14	18.72	31.96	0.424	1517.6	
Flagged	2,120	Utilization	156.0	79.54	128.9	100.0	948.5	0.721
		Claims	160.9	110.4	130.3	1.114	955.5	
Unflagged, > 20, and Non Post Care	73,810	Utilization	33.71	14.34	28.85	20.00	99.99	0.918
		Claims	36.46	15.90	31.64	3.915	311.1	
Unflagged, > 20, and Post Care	7,162	Utilization	42.24	18.61	37.00	20.00	99.84	0.535
		Claims	44.12	36.09	36.78	0.424	1517.6	
Flagged and Non Post Care	1,340	Utilization	143.0	61.15	123.7	100.0	948.5	0.960
		Claims	146.9	61.3	128.5	45.0	955.5	
Flagged and Post Care	780	Utilization	178.2	99.99	140.5	100.0	740.6	0.635
		Claims	184.9	160.6	139.5	1.114	938.5	

Notes: The "Post Care" physicians are those who bill any of the 3 common post service care only codes (hcpcs codes 66984, 66821, and 66982) in the Utilization data. The flagged physicians (greater than 100 hours per week) and unflagged and greater than 20 hours per week are based on the time estimates using the Utilization data and the original service times (same as Fang and Gong (2017)). The claims data total time estimates use the unadjusted service counts. Both the Utilization and claims data total time estimates use the original service time estimates.

involves removing the natural lens.² The median hours per week in the utilization data on these 3 procedures is 125.3 for physicians flagged in both the utilization and claims data, 166.4 for the physicians flagged in the utilization data only, and 60.32 for physicians flagged in the claims data only. Based on the payment the physicians receive, this implies a median gross hourly wage (before paying for supplies, equipment, staff, and other expenses) of about \$1 for physicians flagged in both, \$0.61 for physicians flagged in the utilization data only, and \$1.27 for physicians flagged in the claims data only for these procedures.

C.3 Simulation Evidence

Assessing the performance of the method to estimate total working time is limited by not having 100% of claims. Also, the adjusted service counts are only available for the 5% sample and there is no way to construct adjusted service counts in the Utilization data. The variable of interest in the analysis is the change in flagging propensity that results from the service count adjustment. The following simulation starts by taking the 5% sample as the complete set of data. From this data I calculate a true value of total working time based on adjusted (T^{adj}) and unadjusted service counts (T^{noadj}) as well as the total allowed amount (A) for each of the N physicians. In the total time calculated based on the utilization data, 100 hours corresponds to approximately the 99.66 percentile. In the 5% data using the unadjusted service counts, the corresponding percentile of the total time distribution is 5.766 hours per week. Individuals who work more than 5.766 hours per week are flagged. The true number of flagged physicians is denoted F^{adj} and F^{noadj} depending on whether the total hours calculation uses the adjusted or unadjusted service counts.

Each of 500 simulated samples includes a randomly selected set of 10% of the individual patients who appear in the 5% claims data. Using the claims associated with the individuals in sample i , I calculate total time using both adjusted (t_i^{adj}) and unadjusted service counts (t_i^{noadj}) for each of the n_i physicians who appear in the sample. Following the method in the paper, the inverse of the ratio of allowed charges in the 10% sample is used to estimate total working time for each physician. Some physicians will not appear in the sample and results are presented for all physicians with the unsampled physicians having zero estimated working hours and on the set of physicians who appear in both samples (by setting the estimated hours to missing).³ The number of flagged physicians in the sample are denoted f_i^{adj} and f_i^{noadj} . In addition to the number of flagged physicians, for each simulation the average \bar{t}_i and standard deviation (σ_i) of the total working times is calculated.

Table 8 reports the results of the simulations. Of the 608,050 total physicians in the claims data, on average approximately 390,000 are present in a particular 10% sample. When setting the missing physicians to zero, the average working time across samples is less than the true average working time and the standard deviation is higher. When the average working time is calculated

²Prior to April 2013, Medicare billing processing did not automatically reject claims for cataract surgery performed on the same eye as a previous cataract surgery, so some of these claims could represent inappropriate repeat procedures (Department of Health and Human Services: Office of Inspector General, 2014).

³By setting the unsampled to missing, the “true” parameter values are calculated based on the sampled individuals and will change depending on the sample.

Table 7: Utilization Data Summary Statistics for Common Post-Service Care Only Procedures

Variable	Mean	SD	Median	Min	Max
<u>All Post-Service Care Physicians, N=15,154</u>					
Total Service Count	378.3	907.8	112	11	14,957
Unique Beneficiaries	120.3	164.8	58	11	2,276
Total Allowed Charges	89,048	131,761	36,245	29.5	1,606,401
Hours per Week	17.17	42.43	4.882	0.211	722.9
Wage	187.7	118.0	250.6	0.424	435.3
<u>Flagged in both claims and Utilization data, N=581</u>					
Total Service Count	3,390	2,645	2,652	13	14,957
Unique Beneficiaries	315	447	57	12	2276
Total Allowed Charges	224,791	347,878	11,050	2,542	1,606,401
Hours per Week	157.9	122.5	125.3	0.250	722.9
Wage	82.3	117.3	0.999	0.424	380.9
<u>Flagged in Utilization but not claims, N=199</u>					
Total Service Count	2,738	1,153	2436	24	9,398
Unique Beneficiaries	69.96	150.7	27	12	1,243
Total Allowed Charges	37,191	116,486	4,980	2,722	863,184
Hours per Week	131.7	56.39	116.4	0.461	453.1
Wage	15.41	55.24	0.611	0.523	323.5
<u>Flagged in claims but not Utilization, N=233</u>					
Total Service Count	1,187	533.4	1,260	18	2,103
Unique Beneficiaries	80.09	154.6	18	11	1,016
Total Allowed Charges	51,275	124,710	3,665	860.2	762,233
Hours per Week	56.35	25.85	60.32	0.346	98.31
Wage	52.06	105.6	1.265	0.526	393.0
<u>Flagged in neither and > 20 hrs in Utilization, N=6,932</u>					
Total Service Count	363.6	342.0	256	11	2,291
Unique Beneficiaries	191.3	157.9	156	11	1,428
Total Allowed Charges	149,972	126,674	124,270	68.18	1,019,614
Hours per Week	15.95	15.83	10.99	0.211	98.41
Wage	240.0	89.11	264.3	0.501	435.33

Notes: Physician totals for the 3 common post-service care only codes (hcpcs 66984, 66821, and 66982) in the Utilization data. Physicians are flagged based on a 100 hr per week threshold either the claims or Utilization data using the original time estimates.

Table 8: Simulation Results

Variable	Description	True Value	Mean	SD
<u>All Physicians:</u>				
n_i	# of physicians in the sample	-	390258.0	540.8
\bar{t}_i^{unadj}	Avg total time, unadjusted service counts	0.601	0.545	0.004
σ_i^{unadj}	SD total time, unadjusted service counts	0.936	1.631	2.669
\bar{t}_i^{adj}	Avg total time, adjusted service counts	0.563	0.519	0.004
σ_i^{adj}	SD total time, adjusted service counts	0.745	1.307	2.705
\bar{a}_i/\bar{A}	Avg allowed Ratio	-	0.0996	0.0004
f_i^{noadj}	# flagged using unadjusted service counts	1965	2975.2	39.30
f_i^{adj}	# flagged using adjusted service counts	1014	2460.6	36.62
$f_i^{noadj} - f_i^{adj}$	decrease in flagged physicians	951	514.6	20.89
<u>Only Sampled Physicians:</u>				
\bar{t}_i^{unadj}	Avg total time, unadjusted service counts	0.769 (0.001)	0.808	0.006
σ_i^{unadj}	SD total time, unadjusted service counts	1.048 (0.002)	1.918	3.256
\bar{t}_i^{adj}	Avg total time, adjusted service counts	0.722 (0.001)	0.768	0.006
σ_i^{adj}	SD total time, adjusted service counts	0.830 (0.001)	1.509	3.302
\bar{a}_i/\bar{A}	Avg allowed Ratio	-	0.147	0.0004

Notes: Mean and SD columns are calculated over the values for 500 total simulations. Average and SD in the variable description refer to the statistics calculated over individuals in a simulated sample. The top panel calculates the statistics on all physicians in the 5% claims where unsampled physicians have an estimated time of zero. The bottom panel calculates the statistics only on the physicians who appear in the sample.

on individuals who appear in both the complete data and the sample, the average working time tends to be biased upward.⁴ The higher variance in the total time estimates causes a significantly higher number of physicians to be flagged. The greater noise in the estimated total time compared to the true total times causes a much smaller decrease compared to the true value in the number of flagged physicians when using the adjusted service counts instead of the unadjusted ones. On average, one would calculate a decrease of approximately 515 flagged physicians if one only had access to a given 10% sample of claims when the true decrease is 951 physicians.

D Claims Data Summary Statistics

Table 9 presents summary statistics for a set of variables at the physician level among the claims in the 5% beneficiary sample. The variables include the total Medicare payment to the physician, the total allowed charges, total number of claims and claim lines, total service counts, total number of unique beneficiaries, information about the average work RVU for the physician and the ratio of the total allowed charges in the sample to the 100% allowed charges from the Utilization data summary table. The summary statistics are presented for the physicians who were flagged in the Utilization data as well as those who were not flagged but were estimated to work more than 20

⁴This corresponds to what I do in the main analysis where the sample is restricted to those physicians who appear in both the Utilization data and the 5% Claims data.

Table 9: Claims Data Summary Statistics

Variable	Mean	SD	Median	Min	Max
Unflagged and > 20 hrs/week in the Utilization data, N=80,972					
Total Medicare Payment	18,074	22,289	12842	0	625,845
Total Allowed Charges	23,712	28,540	17,027	24.3	793,296
Number of Claims	142	102	118	1	4105
Number of Claim Lines	245	217	194	1	9499
Total Service Count	459	1,598	221	1	109,717
Number of Beneficiaries	57.5	53.7	43	1	2160
Average wRVU (equal weight by claim line)	1.53	1.52	1.26	0.01	37.58
Average wRVU (service count weighted)	1.44	1.56	1.15	0.004	37.58
Medicare allowed ratio	0.050	0.014	0.049	1.6e-4	0.257
Flagged (> 100 hrs/week) in the Utilization data, N=2,120					
Total Medicare Payment	55,930	69,279	42,212	257	965,091
Total Allowed Charges	72,798	88,533	55,829	328	1,226,518
Number of Claims	232	362	185	4	14,000
Number of Claim Lines	498	1,229	386	4	53,018
Total Service Count	768	1,614	534	4	54,606
Number of Beneficiaries	80.3	101	60	1	2,879
Average wRVU (equal weight by claim line)	1.76	1.16	1.42	0.01	19.1
Average wRVU (service count weighted)	2.62	2.43	1.50	0.02	19.1
Medicare allowed ratio	0.0499	0.011	0.0495	0.0044	0.0962

Notes: Total values by physician for claims in the 5% beneficiary sample. The medicare allowed ratio is defined as the total allowed charges in the 5% sample divided by the 100% allowed charges from the Utilization data provider summary table.

hours per week.

E Utilization Data Results

In this section, I present the results for the number of physicians flagged in the Utilization data using the original time estimate, the time estimates with corrected methodology, and the time estimates with both corrected methodology and new data.

Table 10 reproduces Table 2 in Fang and Gong (2017). The numbers in the table are not identical to those reported in Fang and Gong (2017) because there is an error in the authors’ code that generates the table, which I fixed.⁵ Table 11 presents the results using the original data with the corrected methodology, and Table 12 reports the results using the new data and corrected methodology. Most of the decrease in the number of flagged physicians at the 100 hour per week threshold is due to correcting the methodology, although there is still a substantial decrease in the number of flagged physicians when using the objective time measures from the final Urban Institute study report (a decrease of 296 in 2012 and 238 in 2013).

⁵The authors intended to restrict the summary stats for unflagged physicians to those who work greater than 20 hours in either 2012 or 2013. In the authors code that generates table 2, this is implemented using the restriction “if [flag]==0 & hours2012>20 | hours2013>20”. Parentheses must be added around the “or” statement to generate the sample restriction the authors describe in the paper. This error only has a relatively minor impact on reported

Table 10: Number of Physicians Flagged, Original Fang and Gong (2017) Table

Hours threshold:	<u>80+</u>		<u>100+</u>		<u>112+</u>		<u>168+</u>	
	2012	2013	2012	2013	2012	2013	2012	2013
# of physicians flagged	4125	3838	2292	2120	1689	1546	615	530
Percent of physicians working 20+ hrs/week	4.879	4.618	2.711	2.551	1.998	1.860	0.727	0.638
Percent of all physicians	0.661	0.615	0.367	0.340	0.271	0.248	0.099	0.085
Percent of zero-timed codes (flagged)	9.008	9.147	7.352	7.377	6.430	6.683	4.141	4.250
Percent of zero-timed codes (unflagged)	13.32	13.47	13.28	13.43	13.26	13.41	13.19	13.35
Percent of wRVU from zero-timed codes (flagged)	0.171	0.172	0.115	0.108	0.107	0.098	0.033	0.054
Percent of wRVU from zero-timed codes (unflagged)	0.608	0.629	0.601	0.622	0.598	0.619	0.593	0.614
Percent of volume from zero-timed codes (flagged)	11.43	11.85	9.173	9.391	8.152	8.134	4.104	4.421
Percent of volume from zero-timed codes (unflagged)	15.85	16.00	15.82	15.98	15.79	15.96	15.73	15.90
Percent of revenue from zero-timed codes (flagged)	8.112	9.306	7.106	8.025	6.372	7.018	3.557	3.772
Percent of revenue from zero-timed codes (unflagged)	7.389	7.678	7.428	7.737	7.439	7.755	7.445	7.765
Total # of physicians working 20+ hr/week					96033			
Total # of physicians					623959			

Notes: The hours threshold indicates the number (or fraction) of physicians working at least that many hours per week. Zero-timed codes are code for which there either is no time estimate or an estimate of zero. Unflagged physician is restricted to the set that work more than 20 hours per week in either year.

Table 11: Number of Physicians Flagged, Corrected Methodology

Hours threshold:	<u>80+</u>		<u>100+</u>		<u>112+</u>		<u>168+</u>	
	2012	2013	2012	2013	2012	2013	2012	2013
# of physicians flagged	3586	3260	1960	1756	1444	1292	539	453
Percent of physicians working 20+ hrs/week	4.550	4.233	2.487	2.280	1.832	1.678	0.684	0.588
Percent of all physicians	0.575	0.522	0.314	0.281	0.231	0.207	0.086	0.073
Percent of zero-timed codes (flagged)	9.086	9.086	6.755	6.927	5.899	6.070	3.386	3.539
Percent of zero-timed codes (unflagged)	14.89	15.07	14.83	15.01	14.80	14.98	14.72	14.91
Percent of wRVU from zero-timed codes (flagged)	0.241	0.251	0.177	0.164	0.157	0.139	0.073	0.032
Percent of wRVU from zero-timed codes (unflagged)	0.856	0.871	0.846	0.862	0.843	0.859	0.836	0.853
Percent of volume from zero-timed codes (flagged)	11.01	11.01	8.304	8.149	7.167	7.134	3.299	3.533
Percent of volume from zero-timed codes (unflagged)	16.87	17.05	16.83	17.00	16.79	16.97	16.72	16.90
Percent of revenue from zero-timed codes (flagged)	7.152	7.912	5.598	6.114	5.036	5.551	2.460	2.894
Percent of revenue from zero-timed codes (unflagged)	8.163	8.486	8.179	8.512	8.173	8.508	8.157	8.493
Total # of physicians working 20+ hr/week					89047			
Total # of physicians					623959			

*See notes for Table 10.

Table 12: Number of Physicians Flagged, Corrected Methodology and Zuckerman (2016) data

Hours threshold:	<u>80+</u>		<u>100+</u>		<u>112+</u>		<u>168+</u>	
	2012	2013	2012	2013	2012	2013	2012	2013
# of physicians flagged	2949	2705	1664	1518	1240	1101	446	391
Percent of physicians working 20+ hrs/week	4.473	4.214	2.524	2.365	1.881	1.715	0.677	0.609
Percent of all physicians	0.473	0.434	0.267	0.243	0.199	0.176	0.071	0.063
Percent of zero-timed codes (flagged)	11.18	11.00	8.502	8.717	7.628	7.518	5.842	5.982
Percent of zero-timed codes (unflagged)	22.23	22.40	22.09	22.26	22.03	22.20	21.89	22.07
Percent of wRVU from zero-timed codes (flagged)	0.549	0.514	0.379	0.369	0.327	0.295	0.170	0.151
Percent of wRVU from zero-timed codes (unflagged)	2.117	2.153	2.093	2.129	2.084	2.121	2.066	2.104
Percent of volume from zero-timed codes (flagged)	9.703	9.184	6.865	6.626	5.683	5.293	2.894	3.098
Percent of volume from zero-timed codes (unflagged)	23.07	23.40	22.90	23.22	22.83	23.15	22.66	22.99
Percent of revenue from zero-timed codes (flagged)	4.387	4.375	3.117	3.099	2.611	2.703	1.564	1.543
Percent of revenue from zero-timed codes (unflagged)	9.579	9.991	9.516	9.927	9.488	9.894	9.421	9.831
Total # of physicians working 20+ hr/week	74668							
Total # of physicians	623959							

See notes for Table 10.

E.1 Using Intra-Service Time to Estimate Time Worked

This section calculates physician time using intra-service time instead of total time using the Utilization data. Some pre- and post-service care can be provided by employees of the physician (nurses, physician assistants, etc.), so the intra-service time represents the lower bound of total physician time for a procedure (in most cases). Additionally, objective time measures are only for intra-service time for the timed codes from the Urban Institute study. The first rows present the results using the original data and methodology. At the 100 hour per week threshold, only about a quarter of the number of physicians are flagged compared to using total service time in the calculation of working time. The number of physicians flagged when using intra-service time with the corrected methodology and new data is lower by 91 in 2012 and 31 in 2013. Since there are also fewer physicians estimated to be working more than 20 hours per week, the number of flagged physicians as a percentage of those working more than 20 hours per week increases slightly.

E.2 Using Unique Beneficiary Days to Estimate Physician Time

The Utilization data includes a variable for the number of times a procedure is performed on a unique patient on a given day. This likely overcorrects the service count issue because many services can be legitimately performed multiple times per day or with a service count of more than 1. Using this measure also undercorrects the service count in situations when the physician provides less than the full procedure or when it is conducted at the same time as other procedures (multiple procedure adjustment). Table 14 presents the results using this alternative measure of services in the Utilization data.

values and explains why the unflagged physician values reported in the paper change little across hours thresholds.

Table 13: Flagged Physicians using Intra-Service Time

Hours threshold	<u>80+</u>		<u>100+</u>		<u>112+</u>		<u>168+</u>	
	2012	2013	2012	2013	2012	2013	2012	2013
Time Estimates Based on Original Data:								
# of physicians flagged	1386	1199	574	490	344	322	74	72
% of physicians working 20+ hrs/week	2.377	2.127	0.985	0.869	0.590	0.571	0.127	0.128
# working 20+ hrs/week (either year)	66126							
Time Estimates Based on Corrected Methodology and New Data:								
# of physicians flagged	1137	1052	483	461	305	285	63	60
% of physicians working 20+ hrs/week	2.401	2.335	1.020	1.023	0.644	0.633	0.133	0.133
# working 20+ hrs/week (either year)	53388							
Notes: Number of physicians flagged at different thresholds using the Utilization data and time estimates for intra-service time instead of total procedure time.								

Table 14: Flagged Physicians using Unique Beneficiary Days

Hours threshold	<u>80+</u>		<u>100+</u>		<u>112+</u>		<u>168+</u>	
	2012	2013	2012	2013	2012	2013	2012	2013
Time Estimates Based on Original Data:								
# of physicians flagged	2593	2501	1210	1162	779	763	164	163
% of physicians working 20+ hrs/week	3.406	3.321	1.589	1.543	1.023	1.013	0.215	0.216
# working 20+ hrs/week (in either year)	86123							
Time Estimates Based on Corrected Methodology and New Data:								
# of physicians flagged	1875	1766	866	866	585	582	135	140
% of physicians working 20+ hrs/week	3.057	2.940	1.412	1.442	0.954	0.969	0.220	0.233
# working 20+ hrs/week (in either year)	69283							
Notes: Number of physicians flagged at different thresholds using the Utilization data and the unique beneficiary day measure for the number of services.								

F Analysis of Remaining Problem Codes

In this section, I examine whether there are alternative explanations for the implausibly long physicians hours estimated even after using more conservative service time estimates and attempting to adjust service counts to correct for the overcounting of services. The key assumption in this analysis is that the time component of the work RVU only includes the billing physician’s time. The authors are aware that this assumption is fundamental to their analysis and discuss it in depth in section A of the Online Appendix.⁶ They conclude that the exceptions to the billing provider not being the person who performs the service will have minimal influence. Even if the aggregate effect is minimal, the effect on the relatively small number of flagged physicians could be substantial.

Services by a non-physician practitioner that is employed by the physician are said to be performed “incident to” the physician’s services if they are part of the normal course of treatment and are performed under the supervision of the physician. The physician does not have to be in the room when the “incident to” services are being performed, but the physician must be immediately available if the non-physician practitioner requires assistance. In a non-institutional setting, the physician must be in the office suite, although if the physician is part of a group practice it is only required that any physician be available to assist if necessary.⁷ The authors make the claim that the requirements for “incident to” services mean “the physician is spending almost the same amount of time as the case where she herself furnishes the service” (see Fang and Gong (2017) Online Appendix page 2). The “incident to” requirement does not preclude the physician from seeing other patients or doing work on other cases while the “incident to” service is being performed. The only restriction is that the physician would not be able to perform a procedure that could not be interrupted. Therefore, the general guidelines are not as restrictive as the authors indicate.

In addition, there are exceptions to the general “incident to” requirements and situations where the billing guidelines explicitly state that the billing physician need not perform the entire procedure. One exception to the “incident to” requirements is for an evaluation and management (E/M) service split between a physician and a non-physician practitioner in a facility setting (inpatient, outpatient, or emergency).⁸ For split E/M service in a facility setting, the service can be billed under the physician’s NPI as long as there is any face-to-face interaction between the physician and the patient (Centers for Medicare and Medicaid (2017) Ch 12). Also, diagnostic tests have their own “incident to” requirements which can be more or less stringent depending on the particular test. Another situation where a physician can bill for work performed by another is in group practices with multiple physicians with the same specialty. The Medicare Claims Processing Manual states

⁶See also, Fang and Gong (2017) page 565 footnote 5: “It is important for our analysis that all claims under the same NPI are services furnished by the same individual.”

⁷See Centers for Medicare and Medicaid Services (2017) section 60.1B: “Direct supervision in the office setting does not mean that the physician must be present in the same room with his or her aide. However, the physician must be present in the office suite and immediately available to provide assistance and direction throughout the time the aide is performing services.”

⁸The authors note that the “incident to” requirement only applies in non-facility settings, which is generally (but not universally) true because the “incident to” services performed by the facility staff are considered part of the bundled payment to the facility.

that “physicians in the same group practice who are in the same specialty must bill and be paid as though they were a single physician” for E/M services. The issue of split services among physicians in a group practice also arises for surgical procedures. If multiple physicians in the same group practice collectively perform a full surgical package, they are not required to submit separate bills. In these cases, the group can submit a single claim where the physician who performed the surgery is listed as the performing physician on the claim (Centers for Medicare and Medicaid (2017) Ch 12).⁹ For dialysis services, physicians are paid a monthly capitated rate based on the number of visits. The physician who establishes the plan of care and provides ongoing management is the one who bills the service. The actual visits with the patient can be performed by a different physician or a non-physician practitioner. There is no requirement that the billing physician even be present at the facility (Centers for Medicare and Medicaid (2017) Ch8).

Of the physicians who still have an estimated time worked of 100 hours per week using the new procedure times and adjusted service counts, a number of CPT codes were commonly billed and responsible for a large number of hours worked. Table 15 presents the codes that are billed by at least 50 of the remaining flagged physicians that have the highest average hours worked. The code 90960 is billed by 79 of the remaining flagged physician and on average contributes 51.86 hours per week to the total working time of these physicians. This represents on average 41.7% of the total time worked of these physicians. Of these common and high time taking codes, there are several E/M codes and 2 codes for dialysis services. Parts, or even all, of these services could have been performed by a physician other than the billing physician or by a non-physician practitioner. The remaining common codes are mostly dermatology codes. These dermatology codes have higher estimated times than the AMA RUC times (e.g., 17311 has an estimated total time of 196.35 min and an AMA RUC time of 138 min). The Urban Institute study was only able to collect objective time measures for 2 dermatology codes so the estimated times would have a large standard error.¹⁰

G Timed Codes by Code Group

The Urban Institute study only collected a subset of the codes they had initially planned to collect data for. Table 16 presents the distribution of timed and untimed codes by code group in the original paper and with the new data. Using the new data results in 55 fewer timed codes, but the impact is greater in the surgical codes. For example, for the Integumentary code group (Dermatology), the Urban institute planned to collect objective time measures for 12 codes but were only able to collect data for 2.

⁹This would cause overcounting for the physician who performs the surgery (intra-service only) and undercounting for the physicians providing pre- and post-service care who would not be listed on the claim.

¹⁰See the Appendix G for the breakdown of objectively measured codes by specialty.

Table 15: Common High Time Codes Among the Remaining Flagged Physicians

CPT Code	# Flagged Physicians	Average Hours per Week	Average Time Share	Short Description
90960	79	51.86	0.417	Dialysis services, 4 or more visits per month
92014	80	34.45	0.278	Eye and medical exam, established patient
17311	394	32.83	0.225	Mohs micrograph surgery, first stage
99232	264	20.03	0.165	E/M subsequent inpatient care, typically 25 min per day
99233	239	16.97	0.133	E/M subsequent inpatient care, typically 35 min per day
17312	389	16.07	0.107	Mohs micrograph surgery, additional stages
67028	57	13.56	0.101	Injection of drug into eye
92235	58	12.90	0.090	Examination of retinal blood vessels by ophthalmoscope
99309	73	10.86	0.090	E/M subsequent nursing care facility
90961	75	10.74	0.090	Dialysis services, 2 to 3 visits per month
99223	263	9.998	0.080	Initial inpatient care, typically 70 minutes per day
11100	494	9.253	0.064	Skin biopsy
17000	476	9.235	0.066	Removal of benign skin lesion
99214	592	8.715	0.068	Established patient office visit, typically 25 minutes
17004	348	8.67	0.060	Removal of benign skin lesion, 15 or more

Notes: Codes responsible for the highest average hours per week among the remaining 975 flagged physicians who bill the procedure. # flagged physicians are the number that bill the code at least once in the 5% claims. Codes that are billed by fewer than 50 flagged physicians are excluded. The average time share is the average proportion of total working time spent on that code among the flagged physicians.

Table 16: Timed and Untimed Codes by Code Group

Code Group	HCPCS codes	Zuckerman (2014)		Urban Inst Study (2016)	
		# Timed	# Untimed	# Timed	# Untimed
1 - Anesthesia	00100 - 01999	0	172	0	172
2 - Integumentary System	10021 - 19499	12	308	2	318
3 - Musculoskeletal System	20005 - 29999	20	754	8	766
4 - Respiratory System	30000 - 32999	0	176	0	176
5 - Cardiovascular	33000 - 39599	9	408	8	409
6 - Digestive	40490 - 49999	16	413	15	414
7 - Urinary	50010 - 53899	5	176	5	176
8 - Genital System	54000 - 59899	2	147	2	147
9 - Endocrine System	60000 - 60699	0	19	0	19
10 - Nervous System	61000 - 64999	2	257	1	258
11 - Eye and Auditory	65091 - 69990	6	268	2	272
12 - Radiology	70010 - 79999	18	588	6	600
13 - Pathology and Lab	80047 - 89398	5	864	1	868
14 - Medicine	> 90281 & not E/M	96	582	86	592
15 - E/M	generally 99201-99499	60	10	60	10
All	00100 - 99999	251	5142	196	5197
Surgical Only	10021 - 69990	72	2926	43	2955

*The counts of timed codes include the AMA timed codes.

H List of Urban Institute Objective Times

Tables 17 and 18 list the codes from the final report of the Urban Institute study (Zuckerman et al., 2016). The AMA-RUC times are the times that were included in the interim report (Zuckerman et al., 2014). The Urban institute only collected intra-service times, and the reported value is the median of the objectively measured intra-service times. Urban Institute total times are calculated by assuming no change in pre- and post-service times from the AMA-RUC times.

References

- Centers for Medicare and Medicaid (2017). *Medicare Claims Processing Manual*. <https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/Downloads/clm104c12.pdf>.
- Centers for Medicare and Medicaid Services (2017). *Medicare Benefit Policy Manual*. <https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/Downloads/bp102c15.pdf>.
- Department of Health and Human Services: Office of Inspector General (2014). Medicare paid 22 million in 2012 for potentially inappropriate ophthalmology claims. <https://oig.hhs.gov/oei/reports/oei-04-12-00281.pdf>.
- Fang, H. and Q. Gong (2017). Detecting potential overbilling in medicare reimbursement via hours worked. *American Economic Review* (2), 562–591.
- Zuckerman, S., K. Merrell, R. Berenson, N. McCall, R. Lewis, S. Mitchell, M. Shrestha, and T. Oberlander (2014). Development of a model for the valuation of work relative value units: Objective service time task status report. Urban Institute.
- Zuckerman, S., K. Merrell, R. Berenson, S. Mitchell, D. Upadhyay, and R. Lewis (2016). Collecting empirical physician time data: Piloting an approach for validating work relative value units. Urban Institute.

Table 17: AMA RUC versus Urban Institute Times (in Minutes)

HCPCS	AMA RUC times		Urban Institute times		Δ total time (Urban - AMA)
	intra-service	total	intra-service	total	
Integumentary System Codes:					
11042	15	36	25	46	+10
17110	7	29	15	37	+8
Musculoskeletal System Codes:					
22412	100	287	78	265	-22
23472	140	448	112	420	-28
27130	135	478	87	430	-48
27134	240	617	132	509	-108
27236	90	433	81	424	-9
27244	75	438	71	434	-4
27245	80	443	86	449	+6
27447	124	469	83	428	-41
29827	120	334	81	295	-39
Cardiovascular System Codes:					
33208	60	236	46	222	-14
33249	120	249	42	171	-78
33405	197	768	203	774	+6
33430	232	913	201	882	-31
33518	50	112.6	13	75.6	-37
33519	70	139.8	38	107.8	-32
33533	158	682	214	738	+56
35301	144	431	125	412	-19
Digestive System Codes:					
43235	20	63	5	48	-15
43239	34	84.5	6	56.5	-28
44120	134	611	212	689	+78
44140	150	480	265	595	+115
44143	150	607	203	660	+53
44145	180	615	241	676	+61
44160	120	551	159	590	+39
44204	180	455	142	417	-38
44205	165	428.5	128	391.5	-37
45378	30	75	20	65	-10
45380	51.5	118.5	21	88	-30.5
45385	43	74	22	53	-21
47562	80	228	66	214	-14
47563	90	238	82	230	-8
49505	70	198	60	188	-10

Table 18: AMA RUC versus Urban Institute Times (in Minutes), Continued

HCPCS	AMA RUC times		Urban Institute times		Δ total time (Urban - AMA)
	intra-service	total	intra-service	total	
Urinary Codes:					
50590	60	207	35	182	-25
52000	15	42	14	41	-1
52224	30	79	19	68	-11
52281	20	46	16	42	-4
52601	75	355	56	336	-19
Genital System Codes:					
55700	15	65	13	63	-2
55866	210	512	176	478	-34
Nervous System Codes:					
63047	90	362	50	322	-40
Eye and Auditory Codes:					
66982	33	165	22	154	-11
66984	21	147	18	144	-3
Pathology and Lab Codes:					
70450	17	17	5	5	-12
70551	25	45	8	28	-17
71020	3	5	3	5	0
71250	15	25	16	26	+1
72141	30	30	5	5	-25
77080	5	9	2	6	-3
Radiology Codes:					
88305	25	25	2	2	-23
Medicine Codes:					
92928	76	145	62	131	-14
93010	4	5	0.1	1.1	-3.9
93015	20	26	6	12	-14
93306	20	31.5	5	16.5	-15
93458	45	123	22	100	-23
93459	50	133	48	126	-7