Online Appendix

Detecting Potential Overbilling in Medicare Reimbursement via Hours Worked

Hanming Fang* Qing Gong†

September 6, 2016

In Section A of this online appendix we discuss the institutional background regarding billing and individual provider NPIs. In Section B, we describe the representation of specialties among the timed codes. In Section C we list and examine the codes with the longest estimated time. In Section D we show that our results are not driven by residents who may be working long hours. In Section E we show tables comparing physician characteristics unconditional on Hospital Referral Regions (HRRs). In Section F we show that all of our results are robust to using 112 or 168 weekly hours as the flagging thresholds. In Section G we continue to show the robustness of our results using alternative time estimates, namely using the 25th percentile of time needed, using the minimum time needed, using E/M services only, and using the minimum time needed of E/M services only. In Section H we examine whether results on the comparison of hourly revenues between flagged and unflagged physicians are driven mechanically by the composition of RVUs. In Section I we discuss the comparability of our main sample derived from Medicare Utilization and Payment data with the NAMCS data.

A Multiple Physicians Billing Under the Same NPI

In our analysis, it is important that all claims under the same NPI are services furnished by the same individual. This should be the case per the request of NPI-related regulations. NPI was introduced in 2005 to improve the administration of Medicare, Medicaid, and other health

*Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104, and the NBER. E-mail: hanming.fang@econ.upenn.edu.

†Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104. E-mail: qinggong@sas.upenn.edu.

programs, especially to facilitate electronic data transmission. According to the NPI Final Rule by the Department of Health and Human Services, NPIs are only assigned to "individuals and entities that are licensed and do furnish health care," and stay unchanged in most cases.[1] NPIs with "entity type code" of 1 are individual human beings ("individuals"), and those with "entity type code" of 2 are organizational providers ("organizations"), such as hospitals, clinics, and nursing homes. Individual providers who are members of an organization and the organization they are affiliated with need to have separate NPI numbers (Department of Health and Human Services (2004)). In addition, the NPI Final Rule also requires that "[providers], according to Federal statute and regulations, must be issued their own identification numbers in order to bill and receive payments from Medicare." Hence the providers have to bill for their own or have a billing agency do it on their behalf, but cannot bill under other providers' NPIs. Because of its many advantages, NPIs are commonly used in scholarly articles to track physician activity (Gustafson et al. (2011); Welch et al. (2014)). One paper using the unique physician identification numbers (UPINs), which were established before NPIs, acknowledged that "in some cases, different physicians and loosely affiliated practices bill under the same identifier," and that the new NPIs would have avoided this problem (Pham et al. (2007)).

We are confident that in the vast majority of cases the claims filed under the same NPI are from the same provider. However, there may be exceptions to the above rules. In cases where an auxiliary personnel furnished an "incident-to" service following CMS guidelines,[2] the auxiliary personnel may bill under the NPI of the physician who sets the plan of care (POC). However, these exceptions have minimal influence. This is because (a) CMS guidelines for "incident-to" services require that they must be furnished "under the [billing] physician's direct supervision," which means the billing physician must be in the same designated office area, and *immediately* available to provide assistance and direction. This indicates that the physician is spending almost the same amount of time as the case where she herself furnishes the service; (b) the places-of-service for these "incident-to" services are restricted to non-facilities, which only account for part of our data. It is also possible for physician assistants (PAs) in some specialties to provide part of or even all services to certain patients, although it is not clear under whose NPI these services are billed in practice. But CMS requires PAs bill under their own NPIs and receive 85 percent of the amount a physician would receive (American Academy of Physician Assistants, 2010). Note that

[1] An NPI is "a permanent identifier, assigned for life, unless circumstances justify deactivation, such as a health care provider who finds that his or her NPI has been used fraudulently by another entity" (see Department of Health and Human Services (2004)).

[2] CMS defines "incident-to" services as "those services that are furnished incident to physician professional services in the physician's office (whether located in a separate office suite or within an institution) or in a patient's home" (see Centers for Medicare and Medicaid Services (2002)).

Specialty	Num. of codes	% in timed codes*
Integumentary system	12	10.62
Musculoskeletal system	20	17.70
Cardiovascular/hemic/lymphatic/mediastinum	9	7.96
Digestive system	17	15.04
Urinary system	5	4.42
Genital system	2	1.77
Nervous system	2	1.77
Eye/ocular adnexa/auditory system	6	5.31
Radiology	18	15.93
Pathology and lab	5	4.42
Medicine	17	15.04

Table A1: Distribution of specialties among timed codes with objectively measured time

NOTES: * Timed codes refer to codes with objectively measured time from Zuckerman et al. (2014). The 15 code specialties are defined following the AMA coding guideline Gabbert et al. (2012). E/M services and three other specialties, anesthesia, respiratory system, and endocrine system services, are not included. As a result all HCPCS codes from these specialties with no representation (except timed E/M codes and therapies) do not have a time estimate and end up being "zero-time" codes.

deliberately billing services furnished by PAs under a physician NPI is inappropriate billing, and is, in fact, another form of overbilling (though to a lesser extent) given that the providers get more reimbursement than what their services actually deserve.

B Representation of Specialties Among the Timed Codes

We show the representation of specialties among the timed codes in this section. Of the two types of timed codes, the 145 codes with time information from the AMA coding guidelines are either E/M services or therapies; the 112 codes with objective measured time from the 2014 CMS Survey (Zuckerman et al. (2014)) cover services from 11 of the 14 non-E/M code specialties. Table A1 lists distribution of codes across the 11 specialties.

C Time-Taking Codes

The time needed for untimed codes is predicted using OLS estimated coefficients, which could give rise to potential problems with codes whose work RVUs are farther away from the mean. For this reason, we list the most time-taking codes in Table A2.[3] Of the approximately 5,000 HCPCS codes, 59 have estimated time needed above 120 minutes and are filed for more than 100 times, by 20 or more unique physicians and at least one flagged physicians.[4]

[3]We thank an anonymous referee for suggesting making this list.

[4]We focus on codes that are the most time-consuming *and* filed not too infrequently because the goal of examining these codes is to avoid incorrectly *flagging* a physician due to overestimation of the time needed. Codes not associated with any flagged physicians do also introduce noise into our analysis but are only of secondary concern.

Apart from a few exceptions, these services are billed by very few flagged physicians. And even among the flagged physicians who billed them, these services typically make up a small fraction of all services: 30 codes have a *maximum* share of less than 10%, 46 have maximum shares less than 20%, and 55 have maximum shares less than 30%. Note that given the typically small number of flagged physicians who have ever filed these codes, the maximum share may very well be affected by extreme values. Two services that make up a significantly larger share of flagged physicians' services are 27447, repair of the knee joint, and 90960, dialysis services for patients 20 years and older. The estimated service time for these two is 157.88 and 130.96 minutes, respectively, both of which are close to or even smaller than the time needed in practice.[5]

The last row of the table reports the union of physicians who have ever filed one of the above time-taking codes, which is 784 flagged physicians and 22,699 physicians (flagged and unflagged combined). Conditional on being a flagged physician *and* having filed at least one of these codes, all time-taking codes combined still only occupy a small fraction of a physician's total estimated time. The 75th percentile among flagged physicians with a non-zero share is 27.52%, the 90th percentile 43.66%, even though the outlier share is over 95%.

Therefore we conclude that the "time-taking" services typically contribute a small share to a physician's total service time, whether individually or combined. While we could not rule out the possibility of overestimating the time needed for some services, such potential biases will have limited impacts on the key findings because of the general unimportance of these services in physicians' practices.

D Are We Flagging Mostly Residents?

Residents are known to working long and continuous hours; and most regulations in recent years restrict resident working hours to no more than 80 hours per week averaged over four weeks (see Wolman et al. (2009)). Therefore extra caution must be exercised when "flagging" residents, for whom it can be perfectly normal (unfortunately) to have extremely long hours. For this reason, we only include in our main sample physicians at least one year out of medical school (i.e. those graduated in or before 2011). However, residency can range from one to seven years depending on the specialty, so it is still possible that some of the physicians graduated in more recent years are

[5]MedlinePlus by U.S. National Library of Medicine states that the knee joint replacement "takes about 2 hours" (see https://www.nlm.nih.gov/medlineplus/ency/article/002974.htm). Time needed for a dialysis depends on the specific program, but both the American Association of Kidney Patients and the National Kidney Foundation report numbers ranging from 3 to 4 hours (see https://www.aakp.org/education/resourcelibrary/dialysis-resources/item/nocturnal-dialysis-offers-better-health-while-you-sleep.html and https://www.kidney.org/atoz/content/dialysisinfo).

HCPCS	Description	wRVU	Minutes	N(code)	Flagged	N(phy)	p75	p90	max
92986	Catheter based repair of left lower heart (aortic) valve	22.85	577.7	1.460	6	87	4.67	11.25	11.25
15734	Muscle flap wound repair at trunk	19.86	299.4	3.360	8	101	16.15	21.29	21.29
93620	Insertion of catheters for recording, pacing, and attempted induction of abnormal rhythm in right upper and lower heart	11.57	292.5	51.01	29	1253	15.37	22.54	36.15
15738	Muscle flap wound repair of leg	19.04	287.0	0.930	9	40	7.85	28.24	28.24
93650	Insertion of catheters for creation of complete heart block	10.49	265.2	5.270	23	387	2.49	3.36	3.72
15732	Muscle flap wound repair at head and neck	16.38	246.9	6.080	25	141	9.34	13.46	18.05
15937	Removal of pressure sore and bone at sacrum in preparation of muscle flap or skin graft	15.14	228.2	0.510	3	27	3.56	3.56	3.56
63090	Partial or complete removal of spine bone with release of spinal cord or nerves	30.93	212.7	0.700	4	28	12.21	14.88	14.88
27134	Revision of thigh bone and hip joint prosthesis	30.28	210.5	2.790	3	205	1.59	1.59	1.59
93461	Insertion of catheter in right and left heart for imaging of blood vessels or grafts and left lower heart	8.100	204.8	4.200	24	351	1.67	2.32	4.38
77435	Stereotactic radiation treatment management of 1 or more lesions using imaging guidance	11.87	195	9.340	7	416	2.12	4.31	4.31
93460	Insertion of catheter in right and left heart for imaging of blood vessels or grafts and left lower heart	7.350	192.1	55.50	88	2952	2.78	4.59	11.79
22633	Fusion of lower spine bones with removal of disc, combined posterior, posterolateral or interbody approach	27.75	191.5	15.78	7	786	16.49	19.97	19.97
14301	Tissue transfer repair of wound (30.1 sq centimeters to 60.0 sq centimeters)	12.65	190.7	12.25	136	366	4.17	9.53	27.71
77778	Application of radiation source, complex	11.32	186.0	3.670	4	156	5.20	5.66	5.66
27487	Revision of lower thigh bone and both shin bone components of total knee joint prosthesis	27.11	182.8	3.380	6	257	2.40	2.98	2.98
13160	Second repair of surgical wound	12.04	181.5	0.660	3	42	2.12	2.12	2.12
63081	Partial or complete removal of upper spine bone with release of spinal cord or nerves	26.10	179.4	1.880	7	131	2.62	4.33	4.33
15740	Creation of skin and tissue graft	11.80	177.9	1.130	19	37	1.65	5.95	8.86
93613	Insertion of catheters for 3D mapping of electrical impulses to heart muscles	6.990	176.7	30.98	27	1009	8.11	11.98	12.92
15260	Relocation of patient skin to nose, ears, eyelids, and/or lips (20 sq centimeters or less)	11.64	175.4	34.45	330	1136	5.63	8.62	17.97
14061	Tissue transfer repair of wound (10.1 sq centimeters to 30.0 sq centimeters) of eyelids, nose, ears, or lips	11.48	173.0	16.78	249	634	3.95	5.82	21.66
93459	Insertion of catheter in left heart for imaging of blood vessels or grafts and left lower heart	6.600	172.0	93.29	125	4214	3.59	4.97	13.35
22551	Fusion of spine bones with removal of disc at upper spinal column, anterior approach	25	171.0	13.76	6	877	3.06	5.23	5.23
22533	Fusion of lower spine bones with removal of disc, lateral approach	24.79	167.2	0.530	2	39	0.00	0.00	0.00
14041	Tissue transfer repair of wound (10.1 sq centimeters to 30.0 sq centimeters) of the forehead, cheeks, chin, mouth, neck, underarms, genitals, hands, and/or feet	10.83	163.2	26.53	294	833	4.55	8.51	26.62
22612	Fusion of lower spine bones, posterior or posterolateral approach	23.53	160.0	25.70	9	1049	6.62	11.32	11.32
22558	Fusion of spine bones with removal of disc at lower spinal column, anterior approach	23.53	158.7	5.980	5	328	7.92	10.72	10.72
27447	Repair of knee joint	23.25	157.9	136.4	11	2159	20.24	21.10	42.55
93453	Insertion of catheter into right and left heart for diagnosis	6.240	157.8	1.320	5	85	1.57	1.90	1.90
15240	Relocation of patient skin to forehead, cheeks, chin, mouth, neck, underarms, genitals, hands, or feet (20 sq centimeters or less)	10.41	156.9	3.210	112	192	1.54	2.08	6.68

HCPCS	Description	wRVU	Minutes	N(code)	Flagged	N(phy)	p75	p90	max
93456	Insertion of catheter in right heart for X-ray imaging of blood vessels or grafts	6.150	155.5	1.730	3	142	4.14	4.14	4.14
23472	Prosthetic repair of shoulder joint	22.65	153.4	9.420	6	386	3.20	5.43	5.43
15120	Skin graft of face, scalp, eyelids, mouth, neck, ears, eye region, genitals, hands, feet, and/or multiple fingers or toes (first 100 sq cm or less, or 1% body area of infants and children)	10.15	153.0	1.590	21	79	2.45	4.38	7.85
93458	Insertion of catheter in left heart for imaging of blood vessels or grafts and left lower heart	5.850	151.8	486.4	138	8064	11.85	13.81	33.28
95951	Monitoring and localization of seizure activity over 24-hour period using 16-channel electroencephalograph (EEG)	5.990	151.4	34.45	4	260	18.12	21.76	21.76
63056	Release of lower spinal cord or nerves	21.86	150.3	3.940	3	167	7.50	7.50	7.50
93641	Evaluation of single or dual chamber pacing cardioverter-defibrillator and generator at time of implantation or replacement	5.920	149.7	46.39	60	1634	3.25	5.86	10.26
15100	Skin graft at trunk, arms, or legs (first 100 sq cm or less, or 1% body are of infants and children)	9.900	149.2	1.580	17	77	1.40	2.64	2.65
27130	Replacement of thigh bone and hip joint prosthesis	21.79	147.0	67.19	9	1834	5.44	21.78	21.78
14021	Tissue transfer repair of wound (10.1 sq centimeters to 30.0 sq centimeters) of the scalp, arms, or legs	9.720	146.5	9.010	137	290	2.52	5.51	20.31
34802	Repair of bulging (aneurysm) or tear in abdominal aorta	23.79	143.9	4.490	5	356	1.11	1.37	1.37
27486	Revision of one component of total knee joint prosthesis	21.12	142.4	0.900	2	72	1.08	1.08	1.08
22214	Incision of spine to correct deformity at lower spinal column	21.02	141.7	1.350	4	78	2.94	3.15	3.15
15576	Creation of flap graft to eyelids, nose, ears, lips, or mouth	9.370	141.2	1.330	26	65	1.11	3.60	5.79
93455	Insertion of catheter for imaging of heart blood vessels or grafts	5.540	140.1	6.760	19	467	1.62	2.95	3.27
14060	Tissue transfer repair of wound (10 sq centimeters or less) of eyelids, nose, ears, or lips	9.230	139.1	59.98	403	1699	5.84	8.35	19.79
63075	Removal of upper spine disc and release of spinal cord or nerves	19.60	134.8	0.660	3	39	2.31	2.31	2.31
14001	Tissue transfer repair of wound (10.1 sq centimeters to 30.0 sq centimeters) of the trunk	8.780	132.3	3.050	44	115	3.23	4.56	10.96
77301	Management of modulation radiotherapy planning	7.990	131.3	72.48	27	2127	9.51	10.42	13.81
90960	Dialysis services (4 or more physician visits per month), patient 20 years of age and older	5.180	131.0	2033	120	5508	55.99	65.71	95.48
14040	Tissue transfer repair of wound (10 sq centimeters or less) of the forehead, cheeks, chin, mouth, neck, underarms, genitals, hands, or feet	8.600	129.6	44.11	300	1136	5.35	10.27	23.78
63042	Re-exploration of spine repair with release of spinal cord or nerves	18.76	129.0	5.790	5	367	4.96	4.98	4.98
93609	Insertion of catheter for recording to identify origin of abnormal heart rhythm	4.990	126.2	6.480	15	312	2.09	3.57	4.32
63045	Partial removal of upper spine bone with release of spinal cord or nerves	17.95	123.4	1.010	3	100	1.21	1.21	1.21
93642	Evaluation of single or dual chamber pacing cardioverter-defibrillator with programming or reprogramming	4.880	123.4	2.570	8	113	1.73	6.49	6.49
15220	Relocation of patient skin (20 sq centimeters or less) to scalp, arms, or legs	8.090	121.9	1.750	61	106	1.11	1.69	6.12
93454	Insertion of catheter for imaging of heart blood vessels or grafts	4.790	121.1	43.10	52	2011	1.93	3.30	14.42
93452	Insertion of catheter into left heart for diagnosis	4.750	120.1	8.460	8	253	5.14	11.03	11.03
Union					784	22,699	27.52	43.66	95.48

Table A2: HCPCS codes with the longest estimated time needed

NOTES: The table lists HCPCS codes with the longest estimated time needed (over 120 minutes, in descending order) that have been filed by more than 100 times in total and by more than 20 physicians and more than 1 flagged physicians. "wRVU" is the work RVU of the code; "Minutes" are estimated timed needed in minutes; "N(code)" is the number of times the code was filed in one year (in thousands); "Flagged" is the number of unique flagged physicians who have filed the code; "N(phy)" is the number of unique physicians who have filed the code. "p75/p90/max" is the 75th percentile, 90th percentile, and the maximum of the HCPCS codes' share in terms of estimated time in a *flagged* physician's total services billed. The 2012 and 2013 values are very similar, therefore only the 2012 ones are shown in the interest of space.

Hours threshold	80+		100+		112+		168+	
Year	2012	2013	2012	2013	2012	2013	2012	2013
Number of physicians flagged	4125	3838	2292	2120	1689	1546	615	530
Number of possible residents	16	16	11	6	9	5	2	3
Possible residents/flagged (%)	0.388	0.417	0.480	0.283	0.533	0.323	0.325	0.566

Table A3: Number and fraction of possible residents flagged

NOTES: The table reports the number and fraction of possible residents flagged in 2012 and 2013. Possible residents are identified by their year of graduation from medical school. Physician i of specialty s us a possible resident if i graduated in or after $2012 - T_s^R$, where T_s^R is one year plus the typical length of residency for specialty s.

residents.

In order to check that we are not flagging mostly residents, we first identify possible residents in our sample and see how many of them are flagged. We mark physician i of specialty s as a possible resident if i graduated in or after the year $(2012 - T_s^R)$, where T_s^R is one year plus the typical length of residency for specialty s. For example, the typical residency for family practice is 3 years, therefore we mark family practice physicians as possible residents if they graduated in or after 2008. We are adding one year in order identify all possible residents - there could be variations in residency lengths, and some institutions require an extra year after residency to focus on research. Moreover, we use 2012 in the formula for finding possible residents so that the remainder of physicians are not residents in *any* year in our sample, which covers both 2012 and 2013.

Table A3 summarizes the number of possible residents flagged in 2012 and 2013 under varying thresholds. Only a handful of flagged physicians are possible residents, ranging from 16 under the 80-hour threshold to only 2 or 3 under the 168-hour threshold. These possible residents make up less than 0.6% of all flagged physicians. This shows that it is not the residents with long hours that are driving our results.

E Comparison of Physician Characteristics Not Conditioning on HRR

The following two tables correspond to those in Section 4.1 and 4.4, respectively, except that HRR fixed effects are not included. All the results remain qualitatively and quantitatively similar.

In Table A4 we compare the characteristics of physicians across the following groups, according to the column headings: (1) all physicians, (2) never flagged, (3) flagged in any year, (4) flagged in 2012, (5) flagged in 2013, (6) flagged only in 2012, (7) flagged in both 2012 and 2013, and (8) flagged only in 2013. Table A4 shows that flagged physicians are slightly more likely to be male, non-MD, more experienced, provide fewer E/M services, work in substantially smaller group practices (if at

all), and have fewer hospital affiliations.

Table A5 shows the differences between flagged and unflagged physicians in terms of volume decomposition that are highly similar with those found when controlling for HRR fixed effects as reported in Table 6 in the main text. In particular, flagged physicians have significantly larger volumes of service, but much lower hourly revenues.

F Robustness of Results to Alternative Flagging Thresholds

In this section, we show the robustness of our main results to the choice of flagging thresholds. One might be concerned, despite our deliberately conservative estimates of hours worked, that the 100-hour-per-week threshold might have caught physicians whose billing truthfully reflects the services they provide. Now we use the two higher thresholds, 112 and 168 hours per week, and show that the main results persist under these thresholds.

F.1 Who Reported Implausibly Long Hours?

Tables A6 and A7 are counterparts to Table 4 of our paper, except that they use the two alternative thresholds, respectively. The flagged physicians are still more likely to be males, less likely to have an MD, slightly more experienced, work in much smaller group practices and have fewer hospital affiliations. These results are highly similar to those obtained using the 100-hour threshold in terms of sign, magnitude, and the level of statistical significance.

F.2 What are the Specialties of Flagged Physicians?

Tables A8 and A9 show that SFIs for the 7 specialties in Table 5 of our paper remain qualitatively unchanged. The 4 specialties that are over-represented among the flagged physicians, optometry, dermatology, ophthalmology and pathology, still have SFIs above 50; the 3 specialties that are under-represented, nephrology, cardiology and internal medicine, still have SFIs below 50, although with slight changes in their rankings. In fact, the discrepancies in SFIs become larger when we use a higher flagging threshold – over-represented specialties get even larger SFIs, and under-represented specialties get even smaller SFIs.

F.3 What Codes Do Flagged Physicians Tend to Bill?

Figures A1 and A2 plot the relationship between HCPCS Code Flag Indices (CFIs) and the probability a code is filed by a flagged physician. Not surprisingly, the non-linearity is preserved under alternative flagging thresholds and becomes stronger when the threshold is higher.

	(1) All	(2) Never	(3) Ever	(4) 2012	(5) 2013	(6) 2012 only	(7) Both	(8) 2013 only
1(Male)	0.857	0.856	0.891	0.896	0.896	0.877	0.904	0.867
	(0.001)	(0.001)	(0.006)	(0.007)	(0.007)	(0.013)	(0.008)	(0.017)
1(MD)	0.838	0.844	0.660	0.667	0.714	0.500	0.741	0.624
	(0.001)	(0.001)	(0.010)	(0.010)	(0.011)	(0.020)	(0.012)	(0.024)
Experience (years)	24.14	24.12	24.69	25.14	24.20	26.15	24.70	22.71
	(0.034)	(0.034)	(0.191)	(0.208)	(0.218)	(0.393)	(0.243)	(0.470)
# providers in group	87.19	88.92	29.97	31.47	29.63	31.01	31.67	23.55
	(0.869)	(0.890)	(2.981)	(3.416)	(3.538)	(5.438)	(4.302)	(5.825)
# hospital affiliations	2.774	2.813	1.495	1.535	1.512	1.445	1.576	1.321
	(0.006)	(0.006)	(0.035)	(0.039)	(0.040)	(0.072)	(0.046)	(0.077)
1(in Medicare)	0.857	0.857	0.874	0.878	0.874	0.872	0.880	0.855
	(0.001)	(0.001)	(0.006)	(0.007)	(0.007)	(0.013)	(0.008)	(0.015)
1(in ERX)	0.463	0.463	0.466	0.483	0.500	0.365	0.535	0.395
	(0.002)	(0.002)	(0.009)	(0.010)	(0.011)	(0.018)	(0.013)	(0.021)
1(in PQRS)	0.396	0.396	0.399	0.404	0.424	0.327	0.439	0.378
	(0.002)	(0.002)	(0.009)	(0.010)	(0.011)	(0.018)	(0.012)	(0.021)
1(in EHR)	0.416	0.417	0.397	0.397	0.394	0.406	0.394	0.395
	(0.002)	(0.002)	(0.009)	(0.010)	(0.011)	(0.019)	(0.012)	(0.021)
Types of codes 2012	22.46	22.43	23.39	24.49	24.58	19.82	26.56	18.66
	(0.053)	(0.053)	(0.373)	(0.431)	(0.439)	(0.678)	(0.536)	(0.650)
Types of codes 2013	22.38	22.35	23.32	24.09	24.96	18.39	26.62	20.02
	(0.052)	(0.053)	(0.368)	(0.423)	(0.443)	(0.598)	(0.538)	(0.695)
Types of E/M codes 2012	6.179	6.228	4.573	4.678	4.551	4.639	4.695	4.120
	(0.014)	(0.015)	(0.076)	(0.086)	(0.085)	(0.166)	(0.100)	(0.163)
Types of E/M codes 2013	6.158	6.207	4.553	4.623	4.593	4.430	4.708	4.252
	(0.014)	(0.014)	(0.076)	(0.085)	(0.086)	(0.161)	(0.099)	(0.169)
N	96,033	93,209	2,824	2,292	2,120	704	1,588	532

Table A4: Characteristics of flagged physicians vs unflagged physicians

NOTES: The table compares the means of physician characteristics across subgroups (standard errors of the mean estimator are reported in parentheses). We restrict the sample to physicians billing at least 20 hours per week in at least one year. "All" refers to all physicians in this sample. "Never" refers to physicians never flagged in any year. "Ever" refers to those flagged in at least one year. "2012" and "2013" refer to those flagged in 2012 and 2013, respectively. "2012 (2013) only" refers to those only flagged in 2012 (2013) but not the other year. "Both" refers to those flagged in both years. Physician experience is imputed from the year of graduation. # providers in group refers to the number of providers in the group practice where the billing physician works at. It is 1 if the billing physician works in a solo practice. The number of hospital affiliations is top coded at 5 in the data. 1(in Medicare) is an indicator that the physician accepts Medicare-approved payment amount. 1(in ERX) is an indicator for participation in the Medicare Electronic Prescribing (eRx) Incentive Program, which encourages eRx. 1(in PQRS) is an indicator for participation in the Medicare Physician Quality Reporting System Incentive Program, which provides financial incentives to physicians who report quality measures. 1(in EHR) is an indicator for participation in the Medicare Electronic Health Record (EHR) Incentive Program, which uses financial incentives to reward the adoption of certified EHR technology.

	Flagged		Unflagged	
Year	2012	2013	2012	2013
Num. of services provided	12,548.683	12,365.218	4,540.285	4,490.308
	(542.911)	(562.219)	(12.505)	(12.658)
Num. of services per patient	4.167	3.704	2.434	2.376
	(0.096)	(0.089)	(0.013)	(0.009)
Num. of services provided per hour	1.651	1.648	2.880	2.897
	(0.055)	(0.048)	(0.007)	(0.008)
Num. of patients	5,126.103	5,297.308	2,429.509	2,424.339
	(325.043)	(347.764)	(6.546)	(6.582)
Num. of patients per day	14.006	14.513	6.638	6.642
	(0.888)	(0.953)	(0.018)	(0.018)
Num. of patients per hour	0.705	0.725	1.577	1.591
	(0.033)	(0.029)	(0.004)	(0.004)
Medicare payment per service ($)	80.208	83.180	74.811	73.381
	(1.773)	(1.834)	(0.197)	(0.198)
Medicare payment per patient ($)	197.804	193.769	150.639	146.120
	(5.382)	(4.222)	(0.466)	(0.422)
Medicare payment per hour ($)	118.541	118.677	162.010	159.035
	(2.107)	(2.033)	(0.248)	(0.246)
N	2,292	2,120	93,741	93,913

Table A5: Volumes of services supplied: flagged vs. unflagged physicians

NOTES: The table compares the volumes of services furnished by physicians with different flag statuses. Standard errors of the mean estimator are reported in parentheses. We restrict the sample to physicians billing at least 20 hours per week in at least one year. "Num. of patients" is an overestimation of the actual number of distinct patients due to data limitation, because it is the physician-level sum of the number of distinct patients for each code the physician billed. Hence a patient receiving more than one type of service will be counted multiple times. "Num. of patients per day" is the average number of patients per day assuming 366 (365) working days in year 2012 (2013). "Per hour" statistics are calculated using the estimated total hours worked of each physician.

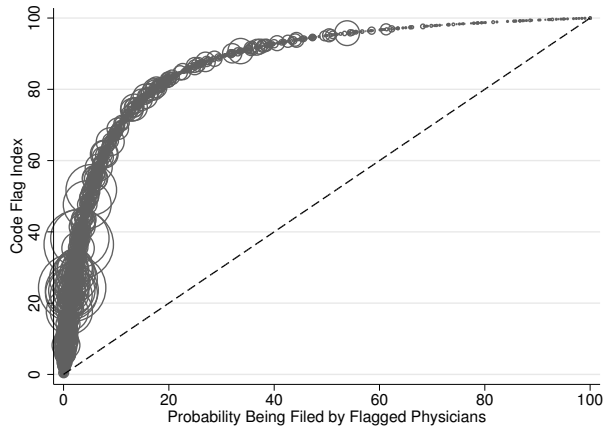


Figure A1: Threshold = 112 hours/week

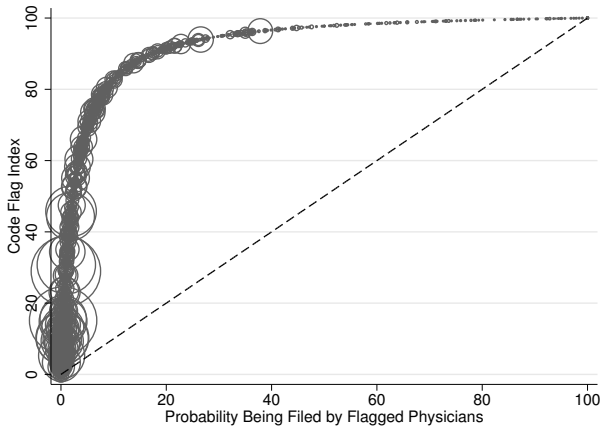


Figure A2: Threshold = 168 hours/week

The Relationship between HCPCS Code Flag Index and its Probability of Being Filed by Flagged Physicians

NOTES: The horizontal axis shows the probability of the HCPCS code being filed by a flagged physician (in %). The vertical axis shows the Code Flag Index (CFI). We restrict the sample to HCPCS codes filed by physicians billing at least 20 hours per week in at least one year. Each circle represents a HCPCS code, with the radius proportional to total revenue. The dashed line is the "45-degree" line.

	(1) Ever	(2) 2012	(3) 2013	(4) 2012 only	(5) Both	(6) 2013 only	Mean of Never
1(male)	0.026***	0.031***	0.034***	0.000	0.045***	-0.001	0.856
	[0.008]	[0.009]	[0.009]	[0.014]	[0.010]	[0.020]	
1(MD)	-0.224***	-0.217***	-0.165***	-0.393***	-0.137***	-0.260***	0.843
	[0.038]	[0.042]	[0.038]	[0.052]	[0.042]	[0.041]	
Experience (years)	0.784***	1.013***	0.420	1.755**	0.648**	-0.299	24.124
	[0.298]	[0.347]	[0.279]	[0.719]	[0.315]	[0.588]	
# providers in group	-53.668***	-53.271***	-52.214***	-57.653***	-51.186***	-55.331***	88.515
	[5.991]	[6.341]	[6.797]	[10.527]	[7.239]	[10.102]	
# hospital affiliations	-1.622***	-1.596***	-1.563***	-1.803***	-1.504***	-1.759***	2.807
	[0.121]	[0.130]	[0.108]	[0.223]	[0.115]	[0.145]	
1(in Medicare)	0.008	0.009	0.006	0.013	0.007	0.001	0.857
	[0.008]	[0.009]	[0.009]	[0.014]	[0.011]	[0.017]	
1(in ERX)	-0.020	-0.011	0.011	-0.113***	0.035	-0.065**	0.463
	[0.018]	[0.020]	[0.020]	[0.026]	[0.022]	[0.029]	
1(in PQRS)	-0.006	-0.007	0.020	-0.081***	0.026	0.001	0.396
	[0.018]	[0.020]	[0.020]	[0.028]	[0.021]	[0.031]	
1(in EHR)	-0.030*	-0.030	-0.044**	0.008	-0.050**	-0.026	0.417
	[0.017]	[0.019]	[0.018]	[0.026]	[0.021]	[0.028]	
Types of codes 2012	-0.207	0.478	1.230	-4.419***	2.708**	-3.631***	22.443
	[1.086]	[1.225]	[1.074]	[1.583]	[1.236]	[1.231]	
Types of codes 2013	-0.101	0.348	1.794*	-5.590***	3.069**	-2.414*	22.366
	[1.086]	[1.214]	[1.086]	[1.418]	[1.240]	[1.277]	
Types of E/M codes 2012	-2.177***	-2.149***	-2.168***	-2.223***	-2.123***	-2.326***	6.223
	[0.183]	[0.192]	[0.181]	[0.314]	[0.190]	[0.233]	
Types of E/M codes 2013	-2.126***	-2.122***	-2.056***	-2.343***	-2.026***	-2.165***	6.201
	[0.183]	[0.191]	[0.182]	[0.303]	[0.190]	[0.243]	
N	2,085	1,689	1,546	539	1,150	396	93,948

Table A6: Characteristics of flagged physicians (threshold being 112 hours/week) vs. unflagged physicians, conditional on Hospital Referral Region (HRR)

NOTES: The table summarizes the difference in physician characteristics between flagged subgroups and the never-flagged subgroup (means reported in the last column) conditional on HRR. We restrict the sample to physicians billing at least 20 hours per week in at least one year. The number in each cell is the estimated coefficient from an OLS regression using the physician characteristic in the corresponding row as the dependent variable, and the flag status dummy (defined by the heading of the column) as the explanatory variable together with HRR fixed effects. Standard errors clustered at the HRR level are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. "All" refers to all physicians in this sample. "Never" refers to physicians never flagged in any year. "Ever" refers to those flagged in at least one year. "2012" and "2013" refer to those flagged in 2012 and 2013, respectively. "2012 (2013) only" refers to those only flagged in 2012 (2013) but not the other year. "Both" refers to those flagged in both years. Physician experience is imputed from the year of graduation. # providers in group refers to the number of providers in the group practice where the billing physician works at. It is 1 if the billing physician does not work in a group practice. The number of hospital affiliations are top coded at 5 in the data. 1(in Medicare) is an indicator that the physician accepts Medicare approved payment amount. 1(in ERX) is an indicator for participation in the Medicare Electronic Prescribing (eRx) Incentive Program, which encourages eRx. 1(in PQRS) is an indicator for participation in the Medicare Physician Quality Reporting System Incentive Program, which provides financial incentives to physicians who report quality measures. 1(in EHR) is an indicator for participation in the Medicare Electronic Health Record (EHR) Incentive Program, which uses financial incentives to reward the adoption of certified EHR technology.

	(1) Ever	(2) 2012	(3) 2013	(4) 2012 only	(5) Both	(6) 2013 only	Mean of Never
1(male)	0.031***	0.043***	0.040***	0.00900	0.064***	-0.0230	0.857
	[0.012]	[0.013]	[0.014]	[0.028]	[0.016]	[0.031]	
1(MD)	-0.380***	-0.366***	-0.324***	-0.512***	-0.281***	-0.447***	0.842
	[0.046]	[0.052]	[0.050]	[0.055]	[0.057]	[0.054]	
Experience (years)	1.081*	1.668***	0.151	3.213***	0.740	-1.410	24.13
	[0.573]	[0.635]	[0.490]	[1.190]	[0.559]	[0.902]	
# providers in group	-55.846***	-52.425***	-58.609***	-49.408*	-54.156***	-70.516***	87.67
	[9.297]	[10.052]	[7.906]	[25.371]	[7.355]	[16.062]	
# hospital affiliations	-2.128***	-2.108***	-1.987***	-2.455***	-1.901***	-2.227***	2.790
	[0.133]	[0.147]	[0.123]	[0.211]	[0.136]	[0.154]	
1(in Medicare)	0.00800	0.00400	0.00200	0.0200	-0.00400	0.0220	0.857
	[0.014]	[0.016]	[0.015]	[0.024]	[0.019]	[0.026]	
1(in ERX)	-0.058**	-0.0410	-0.0370	-0.106***	-0.00100	-0.135***	0.463
	[0.026]	[0.028]	[0.029]	[0.040]	[0.031]	[0.050]	
1(in PQRS)	-0.0140	-0.0150	0.00800	-0.0660	0.0140	-0.00800	0.396
	[0.030]	[0.033]	[0.033]	[0.048]	[0.037]	[0.049]	
1(in EHR)	-0.0310	-0.0480	-0.0310	-0.0300	-0.0590	0.0470	0.416
	[0.031]	[0.034]	[0.034]	[0.048]	[0.038]	[0.044]	
Types of codes 2012	-3.386**	-2.806	-1.088	-8.777***	0.718	-6.070***	22.47
	[1.533]	[1.783]	[1.825]	[1.316]	[2.280]	[1.419]	
Types of codes 2013	-2.865*	-2.532	-0.166	-9.167***	1.394	-4.490***	22.39
	[1.561]	[1.798]	[1.874]	[1.199]	[2.320]	[1.572]	
Types of E/M codes 2012	-2.989***	-3.034***	-2.793***	-3.447***	-2.788***	-2.822***	6.202
	[0.170]	[0.181]	[0.177]	[0.254]	[0.189]	[0.279]	
Types of E/M codes 2013	-2.893***	-2.943***	-2.638***	-3.483***	-2.620***	-2.703***	6.180
	[0.166]	[0.175]	[0.171]	[0.230]	[0.181]	[0.268]	
N	763	615	530	233	382	148	95270

Table A7: Characteristics of flagged physicians (threshold being 168 hours/week) vs. unflagged physicians, conditional on Hospital Referral Region (HRR)
NOTES: See notes to Table A6.

	% in all	Num. unflagged		Num. flagged		SFI	
Specialty\Year		2012	2013	2012	2013	2012	2013
Optometry	1.893	1323	1448	495	370	95.43	93.98
Dermatology	4.185	3661	3638	359	382	84.56	86.52
Ophthalmology	7.960	7379	7383	265	261	66.73	68.36
Pathology	2.746	2585	2587	53	51	53.38	54.65
Nephrology	4.900	4655	4661	51	45	37.96	37.11
Cardiology	11.12	10597	10617	82	62	30.18	26.30
Internal Medicine	11.09	10607	10610	43	40	18.46	18.73
All physicians		94344	94487	1689	1546		

Table A8: Physician specialties and flag status (threshold being 112 hours/week)
NOTES: The table shows seven specialties with the highest SFIs among those with at least 50 flagged physicians. "% in all" shows the fraction of physicians in a specialty among all physicians in our sample (restricted to physicians billing at least 20 hours per week in at least one year). The last row labeled "All physicians" shows the number of flagged (unflagged) physicians by year in our sample.

Specialty\Year	% in all	Num. unflagged		Num. flagged		SFI	
		2012	2013	2012	2013	2012	2013
Optometry	1.893	1,551	1,614	267	204	96.39	95.79
Dermatology	4.185	3,908	3,894	112	126	81.64	85.36
Ophthalmology	7.960	7,551	7,578	93	66	65.65	61.08
Pathology	2.746	2,616	2,617	22	21	56.61	59.12
Nephrology	4.900	4,700	4,698	6	8	16.53	23.48
Internal Medicine	11.09	10,639	10,640	11	10	13.82	14.48
Cardiology	11.12	10,670	10,672	9	7	11.57	10.57
All physicians		95,418	95,503	615	530		

Table A9: Physician specialties and flag status (threshold being 168 hours/week)
NOTES: See notes to Table A8.

Figure A3 and A4 plot the distribution of CFIs under the two alternative thresholds. We still see that high-SFI codes have disproportionately high shares of reimbursement relative to their volumes.

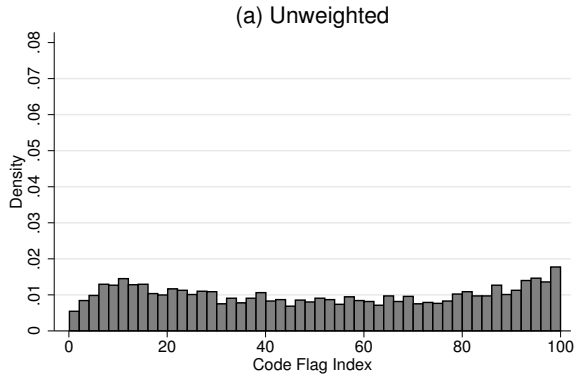
Figures A5 and A6 plot the CFI distributions for codes filed by flagged physicians (solid lines) and by unflagged physicians (dashed lines). Again by construction, flagged physicians file more high-CFI codes. The difference between flagged and unflagged physicians is more dramatic when the higher flagging threshold, 168 hours per week, is used.

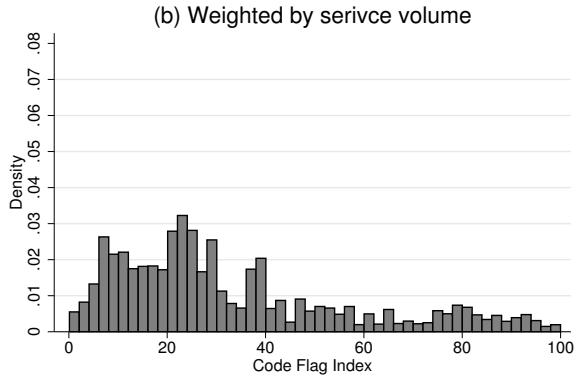
F.4 Decomposing the Long Hours and Quantifying Potential Overbilling

Tables A10 and A11 show how the decomposition of services provided by flagged physicians differs from that of unflagged physicians. Just as Table 8 in our paper shows, flagged physicians provide more services and treat more Medicare Part B FFS patients in total; they also provide more services per patient, and tend to choose services of higher intensity; with average per-service revenues only slightly higher than those of unflagged physicians, they end up with substantially lower imputed hourly revenues. Again, the differences are larger under the 168-hour threshold.

Tables A12 and A13 compare the hourly revenues and Overbilling Potential Factors (OPFs) between flagged and unflagged physicians. The results are still highly similar to those in our paper using the 100-hour threshold, both qualitatively and quantitatively. In particular, flagged physicians have very large discrepancies between their reported and predicted hourly revenues; their OPF1, which captures the excess revenue they get relative to their unflagged peers (assuming identical hours worked), is still around 2; and their OPF2, which describes the extent to which they could be over-reporting hours worked (assuming the goal of overbilling is to achieve the same revenue with fewer hours), ranges between 6.178 and 9.805. The results for unflagged physicians also barely change from those reported in the paper.

Figures A7 and A8 plot the OPFs for both flagged and unflagged physicians using the 112-hour and the 168-hour thresholds, respectively. The distributions of flagged physicians' OPF1 and





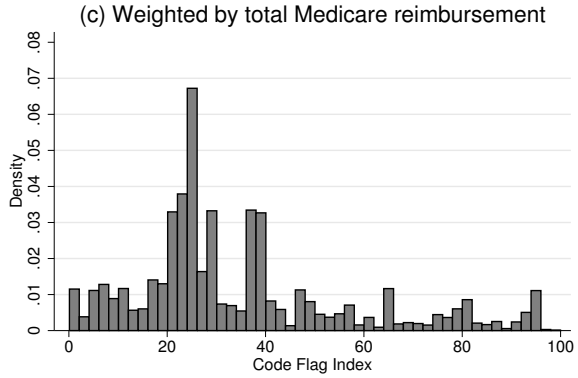
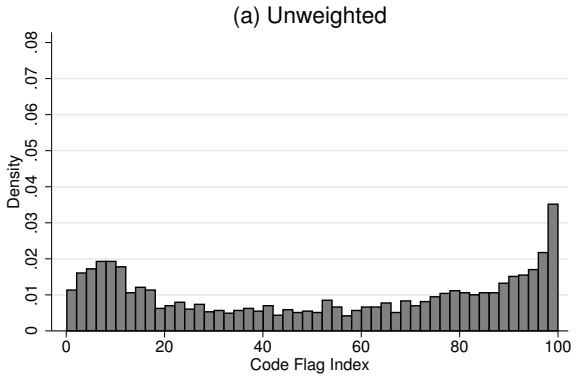
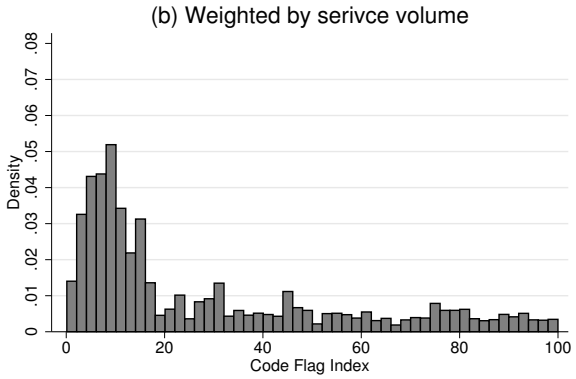


Figure A3: Threshold = 112 hours/week





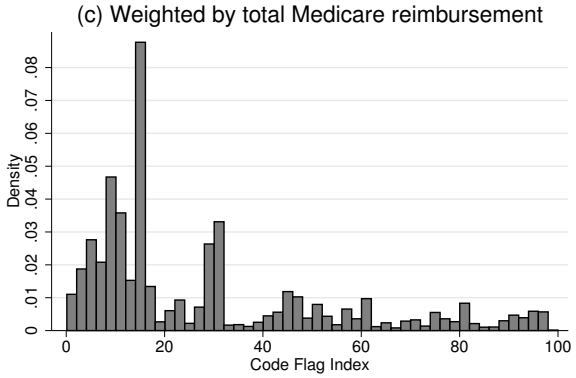


Figure A4: Threshold = 168 hours/week

Distribution of HCPCS Code Flag Index

NOTES: The horizontal axis shows the Code Flag Index (CFI). We restrict the sample to HCPCS codes with CFIs strictly larger than 0 and strictly smaller than 100. Bandwidth is 2 for all three histograms.

	Flagged		Unflagged	
Year	2012	2013	2012	2013
Num. of services provided	8363.404***	8643.937***	4579	4523
	[1047.754]	[1054.903]		
Num. of services per patient	2.007***	1.459***	2.438	2.380
	[0.301]	[0.228]		
Num. of services provided per hour	-1.457***	-1.459***	2.874	2.891
	[0.106]	[0.099]		
Num. of patients	2714.590***	3128.023***	2444	2436
	[597.842]	[627.702]		
Num. of patients per day	7.417***	8.570***	6.677	6.675
	[1.633]	[1.720]		
Num. of patients per hour	-0.998***	-0.984***	1.572	1.586
	[0.069]	[0.063]		
Medicare payment per service ($)	6.011	11.512***	74.91	73.48
	[4.393]	[4.309]		
Medicare payment per patient ($)	44.109***	47.909***	151.1	146.5
	[8.354]	[8.081]		
Medicare payment per hour ($)	-51.834***	-49.186***	161.9	158.9
	[6.336]	[5.439]		
N	1689	1546	94344	94487

Table A10: Volume of services supplied conditional on Hospital Referral Regions: flagged vs. unflagged physicians (threshold being 112 hours/week)

NOTES: The table compares the volume of services furnished by physicians of different subgroups. We restrict the sample to physicians billing at least 20 hours per week in at least one year. The first two columns report the estimation results from OLS regressions using the volume measure in that row as the dependent variable, and the flag dummy as the explanatory variable, together with HRR fixed effects. Standard errors clustered at the HRR level are in brackets. ** $p < 0.05$, *** $p < 0.01$. The last two columns report the means of the two unflagged groups as references. "Num. of patients" is an overestimation of the actual number of distinct patients due to data limitation, because it is the physician-level sum of the number of distinct patients for each code the physician billed. Hence a patient receiving more than one type of service will be counted multiple times. "Num. of patients per day" is the average number of patients per day assuming 366 (365) working days in the year 2012 (2013). "Per hour" statistics are calculated using the estimated total hours worked of each physician.

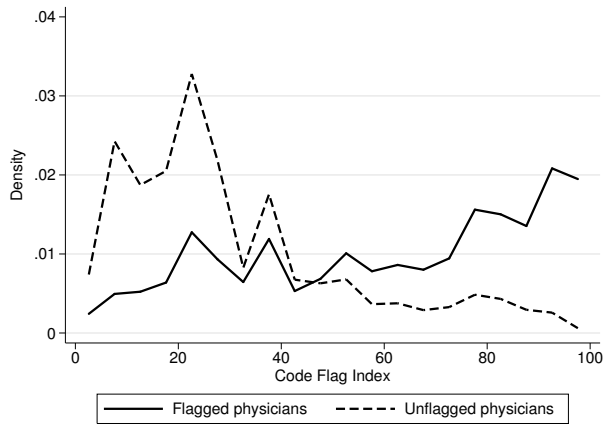


Figure A5: Threshold = 112 hours/week

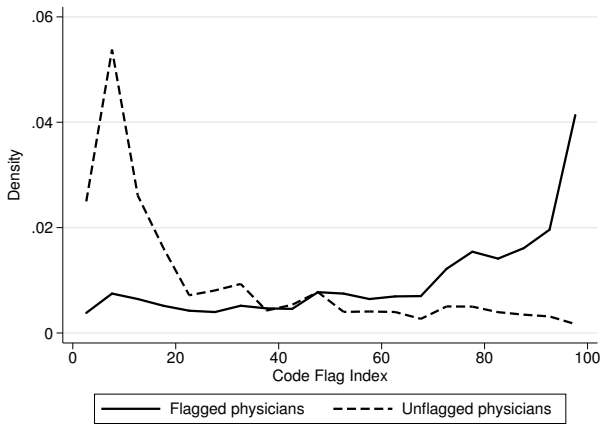


Figure A6: Threshold = 168 hours/week

Distribution of Code Flag Index weighted by service volumes: flagged vs. unflagged physicians

NOTES: The horizontal axis shows the Code Flag Index (CFI). We restrict the sample to HCPCS codes with CFIs strictly larger than 0 and strictly smaller than 100. The solid line shows the distribution of CFIs of codes billed by flagged physicians, and the dashed line shows that for unflagged physicians. Density is weighted by a HCPCS code's total service volume furnished by all physicians.

	Flagged		Unflagged	
Year	2012	2013	2012	2013
Num. of services provided	11979.801***	12488.739***	4654	4596
	[2593.733]	[2728.369]		
Num. of services per patient	3.443***	2.880***	2.453	2.389
	[0.458]	[0.352]		
Num. of services provided per hour	-1.583***	-1.614***	2.861	2.878
	[0.230]	[0.219]		
Num. of patients	3420.234**	4097.479**	2472	2465
	[1485.181]	[1653.190]		
Num. of patients per day	9.345**	11.226**	6.753	6.754
	[4.058]	[4.529]		
Num. of patients per hour	-1.127***	-1.127***	1.563	1.578
	[0.144]	[0.135]		
Medicare payment per service ($)	-13.288***	-7.560	75.06	73.66
	[5.125]	[4.943]		
Medicare payment per patient ($)	26.464**	35.998***	151.7	147.0
	[12.141]	[11.879]		
Medicare payment per hour ($)	-79.931***	-75.784***	161.5	158.6
	[7.233]	[6.974]		
N	615	530	95418	95503

Table A11: Volume of services supplied conditional on Hospital Referral Regions: flagged vs. unflagged physicians (threshold being 168 hours/week)

NOTES: See notes to Table A10.

	Flagged Physicians		Unflagged Physicians	
	2012	2013	2012	2013
Reported hourly revenue ($)	106.909	108.766	158.790	155.861
	(2.418)	(2.357)	(0.243)	(0.241)
Predicted hourly revenue ($)	131.204	134.543	160.605	157.387
	(1.231)	(1.156)	(0.149)	(0.148)
Overbilling Potential Factor 1	1.964	2.031	0.590	0.574
	(0.061)	(0.065)	(0.001)	(0.001)
Overbilling Potential Factor 2	7.347	6.178	1.165	1.143
	(0.277)	(0.270)	(0.005)	(0.004)
N	1,689	1,546	94,344	94,487

Table A12: Hourly revenues and Overbilling Potential Factors (OPFs) (threshold being 112 hours/week)

NOTES: The table compares the hourly revenues and OPFs between flagged and unflagged physicians. We restrict the sample to physicians billing at least 20 hours per week in at least one year. Reported hourly revenues are total revenues divided by total hours reported in one calendar year. Predicted hourly revenues are obtained by first regressing reported hourly revenues on observables (gender, credential, years of experience, a full set of specialty, HRR, and year fixed effects) using the unflagged sample, and then predicting a "fair" hourly revenues for all physicians based on the regression estimates. Standard errors are reported in parentheses.

	Flagged Physicians		Unflagged Physicians	
	2012	2013	2012	2013
Reported hourly revenue ($)	74.501	78.803	158.415	155.527
	(3.927)	(3.962)	(0.243)	(0.241)
Predicted hourly revenue ($)	111.203	117.775	160.179	157.028
	(2.140)	(2.083)	(0.150)	(0.148)
Overbilling Potential Factor 1	2.189	2.315	0.606	0.589
	(0.149)	(0.164)	(0.002)	(0.001)
Overbilling Potential Factor 2	9.805	9.155	1.190	1.163
	(0.449)	(0.505)	(0.005)	(0.004)
N	615	530	95,418	95,503

Table A13: Hourly revenues and Overbilling Potential Factors (OPFs) (threshold being 168 hours/week)

NOTES: See notes to Table A12.

OPF2 are still shifted rightward relative to the distributions of unflagged physicians. Moreover, Panel (b)'s of both figures also show that many flagged physicians' reported revenues fall below their predicted "fair" hourly revenues, whereas the reverse is true for unflagged physicians.

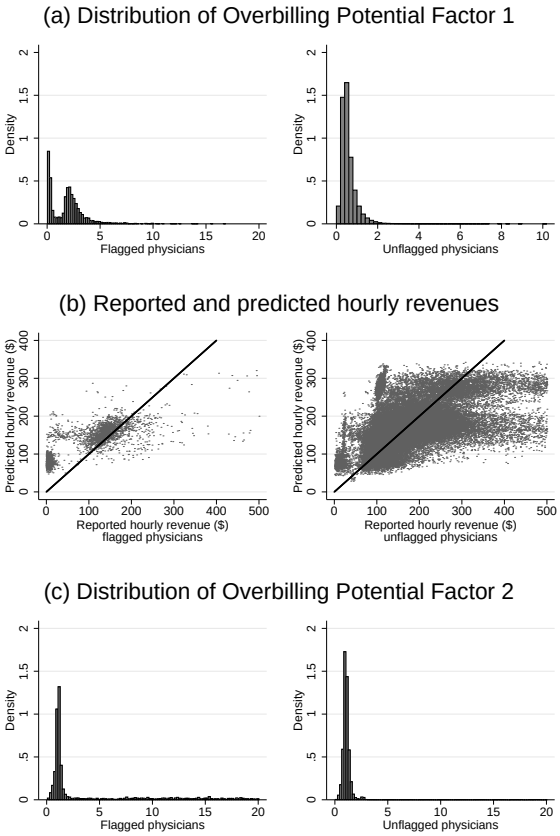


Figure A7: Threshold = 112 hours/week

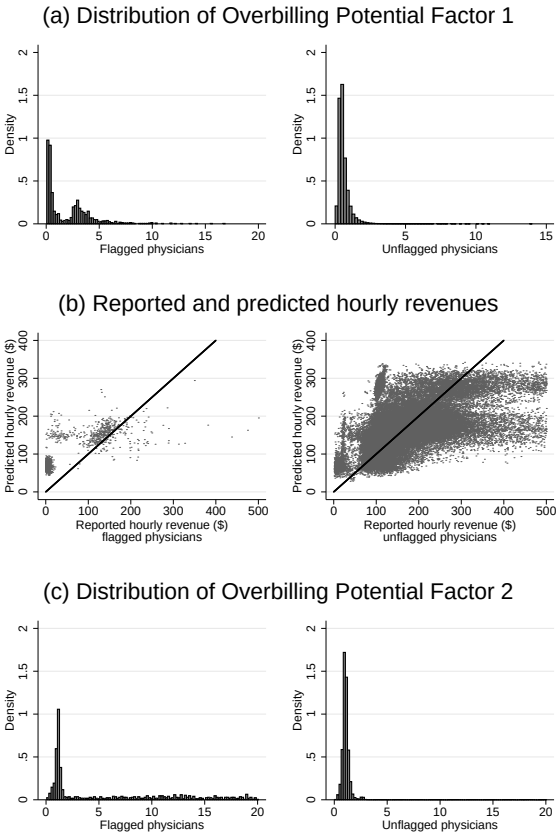


Figure A8: Threshold = 168 hours/week

Overbilling Potential Factors (OPFs)

NOTES: The two figures on the top show the distribution of OPF1 for flagged (left) and unflagged (right) physicians. The two scatter plots in the middle are showing predicted hourly revenues (on the vertical axis, based on OLS regression conditional on physician gender, credential (MD dummy), years of experience, as well as a full set of specialty, HRR, and year fixed effect) against reported hourly revenues (on the horizontal axis). The thick solid line is the "45-degree" line. The two figures on the bottom show the distribution of OPF2 for flagged (left) and unflagged (right) physicians. The bin widths in all four histograms are 0.2.

F.5 Coding Decisions and Fee Differentials

Tables A14 and A15, counterparts to Table 8 of our paper, present the regression results obtained under the two alternative flagging thresholds. All key findings are robust to the choice of thresholds, except that estimates for flag-related variables tend to have larger standard errors. This is because the thresholds used here lead to a much smaller group of flagged physicians, making estimates

noisier.

F.6 Comparison with the Comprehensive Error Rate Testing (CERT) Program

Figures A9 and A10 plot the comparison between HCPCS Code Flag Indices (CFIs) that we constructed using the CMS data and Code Disapproval Indices (CDIs) we calculated using CERT auditing results. Under the higher thresholds (112-hour and 168-hour), CFIs become more extreme, which adds to the incomparability of CFIs and CDIs (see discussion in the paper). This naturally reduces the correlation between the two indices.

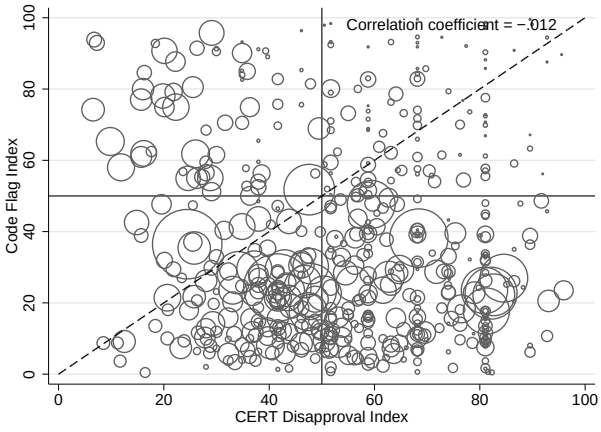


Figure A9: Threshold = 112 hours/week

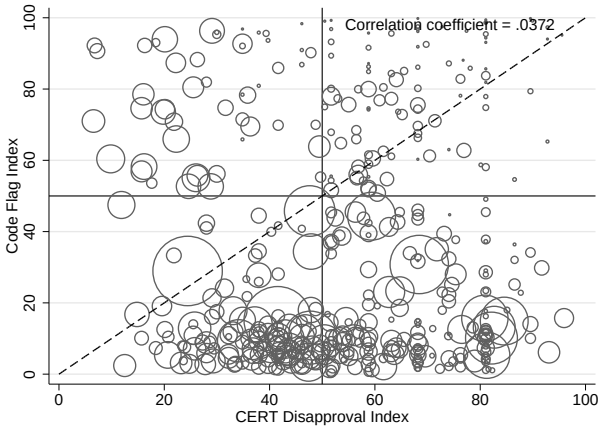


Figure A10: Threshold = 168 hours/week

HCPCS Code Flag Index (CFI) and CERT Code Disapproval Index

NOTES: The horizontal axis shows the CERT Code Disapproval Index. The vertical axis shows the CFI. We restrict the sample to HCPCS codes filed by physicians billing at least 20 hours per week in at least one year and sampled in CERT. The graph has 1621 codes in total. Each circle represents a code, with the radius proportional to total Medicare reimbursement. The dashed line represents cases where the two indices are equal (i.e. a "45-degree" line). The solid horizontal and vertical lines show indices of 50.

	(1)	(2)	(3)	(4)	(5)	(6)
	$K = 3$	$K = 4$	$K = 5$	All K	All K & below average	All K & above average
Flagged	261.5***	653.7*	33.56	178.2***	193.8***	145.9***
	[68.48]	[349.5]	[22.71]	[18.54]	[31.79]	[18.54]
Intensity=2	244.0***	171.9***	12.83***			
	[2.823]	[10.13]	[3.233]			
Intensity=3	130.4***	150.6***	241.2***			
	[2.448]	[10.67]	[3.365]			
Intensity=4		-77.70***	235.9***			
		[10.78]	[3.177]			
Intensity=5			33.69***			
			[3.027]			
Flagged × (intensity=2)	322.2***	-85.40	94.39***			
	[92.27]	[405.6]	[26.17]			
Flagged × (intensity=3)	155.2*	-61.57	151.8***			
	[82.93]	[401.1]	[30.12]			
Flagged × (intensity=4)		395.3	43.17			
		[461.5]	[28.50]			
Flagged × (intensity=5)			14.60			
			[27.22]			
Mid-intensity				240.2***	19.46***	343.9***
				[1.764]	[1.235]	[2.846]
High-intensity				154.0***	34.68***	186.3***
				[1.506]	[1.206]	[2.508]
Flagged × Mid-intensity				83.77***	-86.67***	201.6***
				[26.15]	[31.93]	[36.00]
Flagged × High-intensity				-10.76	-84.58***	71.07**
				[23.46]	[31.37]	[32.40]
HRR	Y	Y	Y	Y	Y	Y
Code cluster	Y	Y	Y	Y	Y	Y
Year	Y	Y	Y	Y	Y	Y
Adjusted R^2	0.190	0.052	0.171	0.157	0.163	0.079
Observations	399,907	53,521	561,657	1,015,085	508,478	506,607

Table A14: Billing patterns and code intensity level (threshold being 112 hours/week)

NOTES: The table reports OLS estimates of the partial effects of code intensity on the number of times the code is filed. We restrict the sample in all specifications to physicians billing at least 20 hours per week in at least one year, and HCPCS codes in the 28 well-defined clusters. Furthermore, Columns (1) to (3) are only using the subsamples of code clusters with 3, 4, and 5 levels of intensities, respectively. Column (4) pools codes in all clusters together, and re-classify intensities to low, middle, and high as specified in our paper. Columns (5) and (6) use the subsample of codes with below- and above-average marginal increases in work RVUs between two adjacent intensity levels, respectively. Physician characteristics, HRR fixed effects, code cluster fixed effects, year fixed effects, and a constant term are included in all specifications but not reported. Standard errors clustered at the physician level are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1)	(2)	(3)	(4)	(5)	(6)
	$K = 3$	$K = 4$	$K = 5$	All K	All K & below average	All K & above average
Flagged	413.1	360.9*	-19.39	158.1***	172.7***	140.6***
	[257.9]	[217.4]	[25.07]	[32.40]	[60.80]	[33.86]
Intensity=2	245.5***	172.2***	15.05***			
	[2.876]	[10.32]	[3.233]			
Intensity=3	131.0***	151.4***	243.3***			
	[2.506]	[10.89]	[3.368]			
Intensity=4		-76.17***	236.3***			
		[10.99]	[3.176]			
Intensity=5			33.35***			
			[3.019]			
Flagged × (intensity=2)	372.8	234.7	133.2***			
	[349.1]	[597.3]	[33.06]			
Flagged × (intensity=3)	299.6	46.25	158.5***			
	[329.9]	[386.2]	[41.95]			
Flagged × (intensity=4)		2161.3**	19.62			
		[942.5]	[35.36]			
Flagged × (intensity=5)			73.31*			
			[42.08]			
Mid-intensity				240.5***	16.75***	346.0***
				[1.763]	[1.045]	[2.852]
High-intensity				152.6***	31.62***	186.2***
				[1.499]	[0.993]	[2.513]
Flagged × Mid-intensity				28.56	-43.80	64.26
				[45.57]	[64.89]	[57.12]
Flagged × High-intensity				-34.08	-40.24	-16.95
				[40.33]	[56.88]	[60.56]
HRR	Y	Y	Y	Y	Y	Y
Code cluster	Y	Y	Y	Y	Y	Y
Year	Y	Y	Y	Y	Y	Y
Adjusted R^2	0.186	0.049	0.170	0.155	0.161	0.077
Observations	399,907	53,521	561,657	1,015,085	508,478	506,607

Table A15: Billing patterns and code intensity level (threshold being 168 hours/week)
NOTES: See notes to Table A14.

G Robustness of Results to Alternative Time Estimates

In this section, we show that our key findings are robust to alternative methods of estimating the hours worked.[6]

Table A16 shows the number and fraction of physicians flagged under the baseline (Panel (a)) and alternative time imputation methods (Panels (b)-(e)). The baseline method is the one used in the main text. We also tried imputing time for all codes using the 25th percentile as well as the minimum in the required time range. For example, if a code requires 30-60 minutes face-to-face with the patient, we set the 25th percentile to be 37.5 minutes and the minimum 30 minutes; if a code typically takes 30 minutes with no minimum requirement, we set the 25th percentile to be 7.5 minutes and the minimum 0. The numbers of flagged physicians, shown in Panels (b) and (c), are about 10% smaller than that in the baseline.

To be even more cautious with imputation, we only use timed codes in Panel (d), and only the minimum time requirement for timed codes in Panel (e). It comes as no surprise that the number of flagged physicians is about 60% smaller than the baseline numbers. But even so, there are still a non-negligible group of physicians billing for long hours. Additionally, the fractions of flagged physicians among those billing more than 20 (imputed) hours per week remain relatively unchanged regardless of the imputation method.

Hours threshold	80+		100+		112+		168+	
Year	2012	2013	2012	2013	2012	2013	2012	2013
(a) Baseline method								
Num. flagged	4125	3838	2292	2120	1689	1546	615	530
% flagged	4.879	4.618	2.711	2.551	1.998	1.860	0.727	0.63
(b) Using the 25th percentile of time needed								
Num. flagged	3686	3443	2097	1937	1581	1457	597	516
% flagged	4.942	4.698	2.812	2.643	2.120	1.988	0.801	0.704
(c) Using the minimum time needed								
Num. flagged	3184	2982	1903	1751	1463	1338	580	490
% flagged	5.285	5.062	3.159	2.972	2.428	2.271	0.963	0.832
(d) Using timed codes only								
Num. flagged	1639	1449	948	826	771	627	362	279
% flagged	3.223	2.902	1.864	1.654	1.516	1.256	0.712	0.559
(e) Using the minimum time needed of timed codes only								
Num. flagged	1188	1043	810	679	693	544	348	268
% flagged	4.243	3.757	2.893	2.446	2.475	1.960	1.243	0.965

Table A16: Number and fraction of physicians flagged under the baseline and alternative methods
NOTES: The table reports the number and fraction of flagged physicians in calendar years 2012 and 2013. "Hours threshold" shows the cutoff number of hours billed per week above which a provider is flagged. "% flagged" shows the fraction of flagged physicians among physicians who billed at least 20 hours per week in the same calendar year.

Now we show, as in Section F, that the key findings persist under these alternative time estimates.

[6]We thank two anonymous referees for suggesting the alternative time estimation methods.

G.1 Who Reported Implausibly Long Hours?

Table A17 shows how the characteristics of flagged physicians flagged differ from those of unflagged physicians under the four alternative time estimation methods. We only report the comparison between the group of physicians flagged in *any* year and the group of unflagged physicians due to limited space and the similarity of comparison results of other groups. The flagged physicians are still more likely to be males, less likely to have an MD, more experienced, work in much smaller group practices and have fewer hospital affiliations. Once again these results are highly similar to those obtained using the baseline time estimates.

G.2 What are the Specialties of Flagged Physicians?

Table A18 shows that SFIs for the 7 specialties in Table 5 of our paper are also qualitatively unchanged. The specialties over-represented (under-represented) among the flagged physicians remain so under each of the four alternative time estimation methods.

G.3 Decomposing the Long Hours and Quantifying Potential Overbilling

Table A19 shows how the decomposition of services provided by flagged physicians differs from that of unflagged physicians. The findings are similar to those under the baseline time estimates: flagged physicians provide more services and treat more patients, provide more services per patient, gravitate towards services of higher intensity, and have lower imputed hourly revenues. One thing to note here is that the comparison of payments are different in the last two columns, where we only use timed codes to estimate hours worked and to flag physicians. This by construction distorts the flagged group to be physicians concentrating on E/M and therapy services, where most of our timed codes are from. These services happen to receive lower reimbursements, which explains the seemingly strange *lower* Medicare payment per service and per patient relative to unflagged physicians. Another reason for the reversal of per-service and per-patient payment comparison results is that the "benchmark" group has changed under the alternative time estimation methods: some physicians who would be flagged under the baseline method are now classified into the "unflagged" group when we only count the timed codes they provided; and this increases the average payment in the unflagged group, making the payment of flagged physicians lower relative to the unflagged.

Table A20 compares the hourly revenues and Overbilling Potential Factors (OPFs) between flagged and unflagged physicians under the four alternative time estimation methods. The results are again very similar to the baseline: the discrepancy between reported and predicted hourly revenues for unflagged physicians persists regardless of the alternative time estimation method

	(1) 25th percentile	(2) Minimum	(3) Time	(4) Timed minimum
1(male)	0.027***	0.025***	0.004	-0.009
	[0.007]	[0.008]	[0.012]	[0.012]
1(MD)	-0.204***	-0.235***	-0.493***	-0.586***
	[0.035]	[0.036]	[0.041]	[0.032]
Experience (years)	0.585**	0.462	1.288***	1.466***
	[0.272]	[0.300]	[0.434]	[0.481]
# providers in group service	-52.506***	-52.589***	-74.698***	-76.472***
	[5.907]	[6.073]	[9.984]	[10.464]
# hospital affiliations	-1.600***	-1.804***	-2.065***	-2.535***
	[0.107]	[0.099]	[0.172]	[0.103]
1(in Medicare)	0.002	0.003	-0.003	0.000
	[0.008]	[0.008]	[0.010]	[0.012]
1(in ERX)	-0.014	-0.0290	-0.141***	-0.174***
	[0.017]	[0.018]	[0.018]	[0.017]
1(in PQRS)	0.000	0.000	-0.069***	-0.069***
	[0.016]	[0.018]	[0.023]	[0.026]
1(in EHR)	-0.030**	-0.032**	-0.005	0.001
	[0.015]	[0.016]	[0.023]	[0.027]
Types of codes 2012	-0.315	-1.426	-9.178***	-11.502***
	[0.949]	[0.981]	[0.934]	[0.779]
Types of codes 2013	-0.181	-1.239	-8.958***	-11.219***
	[0.949]	[0.981]	[0.933]	[0.779]
Types of E/M codes 2012	-2.299***	-2.796***	-2.467***	-3.624***
	[0.148]	[0.118]	[0.302]	[0.140]
Types of E/M codes 2013	-2.252***	-2.748***	-2.441***	-3.583***
	[0.148]	[0.116]	[0.299]	[0.131]
N	2578	2339	1233	1047

Table A17: Characteristics of flagged physicians vs. unflagged physicians using alternative time estimates, conditional on Hospital Referral Region (HRR)

NOTES: The table summarizes the difference in physician characteristics between those ever flagged and those never flagged conditional on HRR. We restrict the sample to physicians billing at least 20 hours per week in at least one year. The number in each cell is the estimated coefficient from an OLS regression using the physician characteristic in the corresponding row as the dependent variable, and the ever-flag indicator as the explanatory variable together with HRR fixed effects. Standard errors clustered at the HRR level are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Physician experience is imputed from the year of graduation. # providers in group refers to the number of providers in the group practice where the billing physician works at. It is 1 if the billing physician works in a solo practice. The number of hospital affiliations is top coded at 5 in the data. 1(in Medicare) is an indicator that the physician accepts Medicare-approved payment amount. 1(in ERX) is an indicator for participation in the Medicare Electronic Prescribing (eRx) Incentive Program, which encourages eRx. 1(in PQRS) is an indicator for participation in the Medicare Physician Quality Reporting System Incentive Program, which provides financial incentives to physicians who report quality measures. 1(in EHR) is an indicator for participation in the Medicare Electronic Health Record (EHR) Incentive Program, which uses financial incentives to reward the adoption of certified EHR technology.

	(a) 25th percentile			(b) Minimum			(c) Timed			(d) Timed minimum		
Specialty	Unfl.	Fl.	SFI	Unfl.	Fl.	SFI	Unfl.	Fl.	SFI	Unfl.	Fl.	SFI
Optometry	1252	566	95.29	1254	564	95.70	1287	531	97.64	1291	527	97.96
Dermatology	3575	442	84.71	3605	412	84.97	3964	53	57.28	3979	38	52.89
Ophthalmology	7268	375	69.80	7287	356	70.73	7536	107	58.75	7541	102	61.39
Pathology	2568	68	54.26	2568	68	56.71	2607	29	52.73	2608	28	55.79
Nephrology	4635	74	41.70	4658	51	35.13	4708	1	2.085	4709	0	0
Cardiology	10576	103	30.38	10618	61	22.13	10650	29	21.45	10673	6	6.198
Internal Medicine	10604	48	16.86	10624	28	11.53	10611	41	27.93	10644	8	8.118
All physicians	93936	2097		94130	1903		95085	948		95223	810	

Table A18: Physician specialties and flag status using alternative time estimates

NOTES: The table shows seven specialties with the highest SFIs among those with at least 50 flagged physicians. Table only shows results for 2012 and suppresses highly similar results for 2013. "Unfl." columns show the number of unflagged physicians; "Fl." columns show the number of flagged physicians. The last row labeled "All physicians" shows the total number of flagged (unflagged) physicians in our sample.

	25th percentile	Minimum	Timed	Timed minimum
Num. of services provided	7541.191***	7392.312***	6374.464***	5606.092***
	[862.321]	[947.670]	[1133.655]	[1232.388]
Num. of services per patient	1.793***	1.967***	3.909***	4.390***
	[0.269]	[0.284]	[0.331]	[0.315]
Num. of services provided per hour	-1.435***	-1.511***	-1.759***	-1.953***
	[0.092]	[0.098]	[0.107]	[0.100]
Num. of patients	2501.919***	2341.628***	638.2	138.0
	[491.338]	[535.972]	[391.841]	[383.723]
Num. of patients per day	6.836***	6.398***	1.744	0.377
	[1.342]	[1.464]	[1.071]	[1.048]
Num. of patients per hour	-0.973***	-1.022***	-1.294***	-1.397***
	[0.060]	[0.063]	[0.053]	[0.045]
Medicare payment per service ($)	9.359**	9.951**	-36.626***	-41.282***
	[4.182]	[4.644]	[2.552]	[2.143]
Medicare payment per patient ($)	51.682***	53.276***	-14.635**	-26.849***
	[9.540]	[10.552]	[6.813]	[5.129]
Medicare payment per hour ($)	-47.409***	-51.140***	-100.455***	-113.915***
	[5.946]	[6.389]	[5.674]	[4.430]
N	2097	1903	948	810

Table A19: Volume of services supplied conditional on Hospital Referral Regions: flagged vs. unflagged physicians using alternative time estimates

NOTES: The table compares the volume of services furnished by flagged physicians with that by unflagged physicians. Table only shows results for 2012 and suppresses highly similar results for 2013. We restrict the sample to physicians billing at least 20 hours per week in at least one year. The numbers reported are estimation results from OLS regressions using the volume measure in that row as the dependent variable, and the flag dummy as the explanatory variable, together with HRR fixed effects. Standard errors clustered at the HRR level are in brackets. ** $p < 0.05$, *** $p < 0.01$. "Num. of patients" is an overestimation of the actual number of distinct patients due to data limitation, because it is the physician-level sum of the number of distinct patients for each code the physician billed. Hence a patient receiving more than one type of service will be counted multiple times. "Num. of patients per day" is the average number of patients per day assuming 366 working days in the year 2012. "Per hour" statistics are calculated using the estimated total hours worked of each physician.

used; flagged physicians still have a much larger OPF1 than their unflagged peers, which reflects the excess revenue the former get (assuming identical hours worked); flagged physicians also have an average OPF2 that is above an order larger than that of unflagged physicians, which shows the extent of potentially over-reported hours (assuming the goal of overbilling is to achieve the same revenue with fewer hours).

	25th percentile		Minimum		Timed		Timed minimum	
	Flagged	Unflagged	Flagged	Unflagged	Flagged	Unflagged	Flagged	Unflagged
Reported hourly revenue ($)	112.1	157.4	108.1	157.4	56.91	157.4	38.49	157.4
	(1.529)	(0.171)	(1.650)	(0.171)	(1.670)	(0.171)	(1.528)	(0.171)
Predicted hourly revenue ($)	136.0	159.1	133.7	159.1	110.9	159.1	101.6	159.1
	(0.728)	(0.105)	(0.775)	(0.104)	(1.050)	(0.104)	(1.012)	(0.104)
Overbilling Potential Factor 1	1.909	0.577	1.907	0.580	1.187	0.601	1.016	0.603
	(0.036)	(0.001)	(0.040)	(0.001)	(0.046)	(0.001)	(0.052)	(0.001)
Overbilling Potential Factor 2	6.457	1.150	6.997	1.150	12.55	1.152	14.48	1.152
	(0.175)	(0.003)	(0.190)	(0.003)	(0.320)	(0.003)	(0.348)	(0.003)
N	4034	188032	3654	188412	1774	190292	1489	190577

Table A20: Hourly revenues and Overbilling Potential Factors (OPFs) using alternative time estimates

NOTES: The table compares the hourly revenues and OPFs between flagged and unflagged physicians. We restrict the sample to physicians billing at least 20 hours per week in at least one year. Reported hourly revenues are total revenues divided by total hours reported in one calendar year. Predicted hourly revenues are obtained by first regressing reported hourly revenues on observables (gender, credential, years of experience, a full set of specialty, HRR, and year fixed effects) using the unflagged sample, and then predicting a "fair" hourly revenues for all physicians based on the regression estimates. Standard errors are reported in parentheses.

G.4 Coding Decisions and Fee Differentials

Tables A21 through A24 are results for regressions of code choices on physician flag status and code intensities. Most of the key findings under the baseline method remain, though we also notice some changes. Note that the choice of a time estimation method leads to changes in the composition of the "flagged" and the "unflagged" groups. For example, using only timed codes to construct implied hours worked tends to flag physicians whose services focus on E/M or therapy codes. Moreover, when we use the minimum time needed, low-intensity codes are often assigned zero time, so only those who file a lot of mid- or high-intensity codes would be flagged. In other words, the *flagged* indicator already conveys much information about the code intensity, which could also explain the less clear coefficients on the interaction terms between *Flagged* and intensity dummies.

H Discussion of Hourly Revenue Comparison Results

In Table 7 of the paper, we compared the reported hourly revenues of flagged and unflagged physicians. One concern is that whether the significantly lower hourly revenues of flagged physicians

	(1)	(2)	(3)	(4)	(5)	(6)
	$K = 3$	$K = 4$	$K = 5$	All K	All K & below average	All K & above average
Flagged	182.9***	220.2*	33.19*	145.9***	135.8***	140.2***
	[49.55]	[125.4]	[19.50]	[13.82]	[21.96]	[16.59]
Intensity=2	244.1***	171.2***	12.20***			
	[2.839]	[10.29]	[3.242]			
Intensity=3	130.1***	149.1***	241.4***			
	[2.472]	[10.80]	[3.377]			
Intensity=4		-80.44***	236.1***			
		[10.88]	[3.188]			
Intensity=5			33.85***			
			[3.037]			
Flagged × (Intensity=2)	244.5***	148.7	91.02***			
	[73.12]	[221.1]	[22.71]			
Flagged × (Intensity=3)	167.0***	288.9	116.7***			
	[63.36]	[221.8]	[25.75]			
Flagged × (Intensity=4)		775.0**	25.93			
		[304.5]	[24.11]			
Flagged × (Intensity=5)			6.847			
			[23.21]			
Mid-intensity				240.5***	18.67***	345.3***
				[1.761]	[1.147]	[2.851]
High-intensity				153.9***	33.79***	187.2***
				[1.502]	[1.111]	[2.509]
Flagged × Mid-intensity				54.36**	-41.81*	114.5***
				[21.23]	[23.19]	[30.00]
Flagged × High-intensity				-7.711	-34.63	30.41
				[17.96]	[21.41]	[27.82]
HRR	Y	Y	Y	Y	Y	Y
Code cluster	Y	Y	Y	Y	Y	Y
Year	Y	Y	Y	Y	Y	Y
Adjusted R^2	0.188	0.050	0.171	0.156	0.162	0.078
Observations	399907	53521	561657	1015085	508478	506607

Table A21: Billing patterns and code intensity level (using the 25th percentile of time needed)

NOTES: The table reports OLS estimates of the partial effects of code intensity on the number of times the code is filed. We restrict the sample in all specifications to physicians billing at least 20 hours per week in at least one year, and HCPCS codes in the 28 well-defined clusters. Furthermore, Columns (1) to (3) are only using the subsamples of code clusters with 3, 4, and 5 levels of intensities, respectively. Column (4) pools codes in all clusters together, and re-classify intensities to low, middle, and high as specified in our paper. Columns (5) and (6) use the subsample of codes with below- and above-average marginal increases in work RVUs between two adjacent intensity levels, respectively. Physician characteristics, HRR fixed effects, code cluster fixed effects, year fixed effects, and a constant term are included in all specifications but not reported. Standard errors clustered at the physician level are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1)	(2)	(3)	(4)	(5)	(6)
	$K = 3$	$K = 4$	$K = 5$	All K	All K & below average	All K & above average
Flagged	175.5***	116.0	32.57	125.6***	86.52***	145.2***
	[67.42]	[126.5]	[20.94]	[12.95]	[18.22]	[17.68]
Intensity=2	245.4***	173.1***	12.39***			
	[2.872]	[10.38]	[3.239]			
Intensity=3	130.8***	150.7***	242.8***			
	[2.498]	[10.90]	[3.381]			
Intensity=4		-78.69***	236.9***			
		[10.92]	[3.189]			
Intensity=5			33.80***			
			[3.033]			
Flagged × (Intensity=2)	90.29	-183.8	95.67***			
	[85.58]	[143.4]	[24.25]			
Flagged × (Intensity=3)	112.2	183.9	67.97**			
	[84.40]	[230.9]	[26.59]			
Flagged × (Intensity=4)		1231.0**	-24.66			
		[497.9]	[25.00]			
Flagged × (Intensity=5)			20.16			
			[25.95]			
Mid-intensity				241.3***	17.05***	348.3***
				[1.766]	[1.049]	[2.858]
High-intensity				153.9***	32.19***	188.9***
				[1.497]	[1.000]	[2.511]
Flagged × Mid-intensity				-8.245	8.870	-23.65
				[19.18]	[20.03]	[27.37]
Flagged × High-intensity				-39.46**	10.64	-64.79**
				[17.66]	[18.26]	[29.01]
HRR	Y	Y	Y	Y	Y	Y
Code cluster	Y	Y	Y	Y	Y	Y
Year	Y	Y	Y	Y	Y	Y
Adjusted R^2	0.186	0.048	0.170	0.155	0.161	0.077
Observations	399907	53521	561657	1015085	508478	506607

Table A22: Billing patterns and code intensity level (using minimum time needed)

NOTES: The table reports OLS estimates of the partial effects of code intensity on the number of times the code is filed. We restrict the sample in all specifications to physicians billing at least 20 hours per week in at least one year, and HCPCS codes in the 28 well-defined clusters. Furthermore, Columns (1) to (3) are only using the subsamples of code clusters with 3, 4, and 5 levels of intensities, respectively. Column (4) pools codes in all clusters together, and re-classify intensities to low, middle, and high as specified in our paper. Columns (5) and (6) use the subsample of codes with below- and above-average marginal increases in work RVUs between two adjacent intensity levels, respectively. Physician characteristics, HRR fixed effects, code cluster fixed effects, year fixed effects, and a constant term are included in all specifications but not reported. Standard errors clustered at the physician level are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1)	(2)	(3)	(4)	(5)	(6)
	$K = 3$	$K = 4$	$K = 5$	All K	All K & below average	All K & above average
Flagged	329.3**	779.0**	4.952	249.7***	443.4***	106.9***
	[136.8]	[341.3]	[28.66]	[39.99]	[75.96]	[32.83]
Intensity=2	243.1***	169.6***	16.39***			
	[2.817]	[10.02]	[3.238]			
Intensity=3	129.8***	149.6***	244.0***			
	[2.450]	[10.63]	[3.367]			
Intensity=4		-78.95***	235.9***			
		[10.68]	[3.172]			
Intensity=5			33.39***			
			[3.020]			
Flagged × (Intensity=2)	876.1***	280.8	10.57			
	[159.9]	[464.3]	[34.83]			
Flagged × (Intensity=3)	435.2***	-42.86	34.27			
	[154.0]	[402.9]	[40.43]			
Flagged × (Intensity=4)		206.8	70.73*			
		[413.3]	[37.69]			
Flagged × (Intensity=5)			33.33			
			[36.66]			
Mid-intensity				240.1***	18.12***	343.4***
				[1.760]	[1.099]	[2.879]
High-intensity				152.4***	32.76***	183.8***
				[1.504]	[1.047]	[2.543]
Flagged × Mid-intensity				59.35	-323.5***	270.8***
				[52.22]	[76.52]	[62.13]
Flagged × High-intensity				28.88	-294.9***	274.2***
				[47.43]	[75.95]	[53.42]
HRR	Y	Y	Y	Y	Y	Y
Code cluster	Y	Y	Y	Y	Y	Y
Year	Y	Y	Y	Y	Y	Y
Adjusted R^2	0.196	0.058	0.170	0.156	0.164	0.079
Observations	399907	53521	561657	1015085	508478	506607

Table A23: Billing patterns and code intensity level (using timed codes only)

NOTES: The table reports OLS estimates of the partial effects of code intensity on the number of times the code is filed. We restrict the sample in all specifications to physicians billing at least 20 hours per week in at least one year, and HCPCS codes in the 28 well-defined clusters. Furthermore, Columns (1) to (3) are only using the subsamples of code clusters with 3, 4, and 5 levels of intensities, respectively. Column (4) pools codes in all clusters together, and re-classify intensities to low, middle, and high as specified in our paper. Columns (5) and (6) use the subsample of codes with below- and above-average marginal increases in work RVUs between two adjacent intensity levels, respectively. Physician characteristics, HRR fixed effects, code cluster fixed effects, year fixed effects, and a constant term are included in all specifications but not reported. Standard errors clustered at the physician level are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1)	(2)	(3)	(4)	(5)	(6)
	$K = 3$	$K = 4$	$K = 5$	All K	All K & below average	All K & above average
Flagged	285.8	341.0	-41.02	52.08*	133.2*	41.67
	[210.3]	[268.9]	[25.12]	[31.58]	[75.49]	[29.21]
Intensity=2	245.5***	172.8***	16.57***			
	[2.879]	[10.37]	[3.237]			
Intensity=3	130.9***	150.9***	245.0***			
	[2.505]	[10.89]	[3.371]			
Intensity=4		-78.47***	236.6***			
		[10.91]	[3.174]			
Intensity=5			33.45***			
			[3.017]			
Flagged × (Intensity=2)	449.4	-362.4	28.63			
	[302.1]	[326.3]	[31.56]			
Flagged × (Intensity=3)	603.2**	41.81	-43.61			
	[289.3]	[384.9]	[35.61]			
Flagged × (Intensity=4)		1762.3**	-21.02			
		[788.9]	[33.35]			
Flagged × (Intensity=5)			86.17**			
			[42.21]			
Mid-intensity				240.8***	16.17***	346.9***
				[1.766]	[1.001]	[2.869]
High-intensity				152.0***	30.75***	185.8***
				[1.499]	[0.938]	[2.528]
Flagged × Mid-intensity				-90.76**	-62.22	-143.5***
				[38.41]	[78.45]	[42.19]
Flagged × High-intensity				22.55	-7.736	10.39
				[36.88]	[74.64]	[51.06]
HRR	Y	Y	Y	Y	Y	Y
Code cluster	Y	Y	Y	Y	Y	Y
Year	Y	Y	Y	Y	Y	Y
Adjusted R^2	0.186	0.049	0.170	0.154	0.161	0.077
Observations	399907	53521	561657	1015085	508478	506607

Table A24: Billing patterns and code intensity level (using minimum time needed of timed codes only)

NOTES: The table reports OLS estimates of the partial effects of code intensity on the number of times the code is filed. We restrict the sample in all specifications to physicians billing at least 20 hours per week in at least one year, and HCPCS codes in the 28 well-defined clusters. Furthermore, Columns (1) to (3) are only using the subsamples of code clusters with 3, 4, and 5 levels of intensities, respectively. Column (4) pools codes in all clusters together, and re-classify intensities to low, middle, and high as specified in our paper. Columns (5) and (6) use the subsample of codes with below- and above-average marginal increases in work RVUs between two adjacent intensity levels, respectively. Physician characteristics, HRR fixed effects, code cluster fixed effects, year fixed effects, and a constant term are included in all specifications but not reported. Standard errors clustered at the physician level are in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

are a mechanical result of the variation in work RVUs and total RVUs across code intensities. This concern arises from the composition of total RVUs, which consists of work RVU (wRVU), practice expense RVU (peRVU), and malpractice RVU (mpRVU). Recall that we imputed the time needed for untimed codes from their wRVUs and a specialty-specific "time-per-wRVU" obtained from timed codes; and that the service fee for each code is (roughly) total RVU multiplied by a constant number, the Conversion Factor (CF). Hence the following holds conditional on specialty:

$$\begin{aligned} \text{HourlyRev} &= \frac{\text{Fee}}{\text{Imputed time}} = \frac{(\text{Total RVU}) \times CF}{\text{wRVU} \times t/\text{wRVU}} \\ &= \text{Const.} \times \frac{\text{wRVU} + \text{peRVU} + \text{mpRVU}}{wRVU} \end{aligned}$$

where everything is in terms of weighted averages over all the services provided by a given physician. Note that if $\frac{wRVU+peRVU+mpRVU}{wRVU}$ decreases with code intensity (i.e. if wRVU increases with intensity but peRVU and mpRVU do not), then there will be a mechanical decline in hourly revenue for physicians who tend to bill high-intensity codes.

Hence we test whether this relationship is driving the significantly lower hourly revenue of flagged physicians relative to unflagged physicians, and present the results in Table A25.

The dependent variable in all four specifications in Table A25 is the share of wRVU in total RVU (totRVU) in percents for each HCPCS code. Columns (1) and (2) show how wRVU/totRVU changes with totRVU, using totRVU as a general measure of code intensity. Column (1) finds that wRVU/totRVU increases by 0.04 percentage points with a unit increase in totRVU. With a standard deviation increase in totRVU (22.93 units), wRVU/totRVU only increase by $0.04 \times 22.93 = 0.9172$ percentage points, which is economically insignificant given that the average wRVU/totRVU is 41.27%. Column (2) controls for specialty fixed-effects because time imputation is specialty-specific, and finds that wRVU/totRVU even slightly declines for HCPCS codes with larger totRVU, although the effect is statistically marginally significant and economically insignificant.

Columns (3) and (4) focus on the 78 E/M codes with non-zero totRVU and a clear cluster structure. Recall that each code cluster has 3 to 5 HCPCS codes representing the same service at different levels of intensity. Column (3) examines how wRVU/totRVU varies with code intensity *within* a cluster by including cluster fixed-effects, assuming a linear relationship between intensity and the dependent variable. We find that wRVU/totRVU increases by 1.81 percentage points per level of intensity on average, which is still small compared with the average of 66.16% for E/M codes. Relaxing the linearity assumption gives similar results: codes with intensity 2, 3, 4, and 5 are 2.59 to 7.39 percentage points higher in wRVU/totRVU relative to the lowest intensity (Intensity=1).

	Y = wRVU/totRVU (%)			
	(1)	(2)	(3)	(4)
	All	All	E/M	E/M
Total RVU	0.0403*	-0.0370*		
	[0.0241]	[0.0190]		
Code intensity (1 to 5)			1.807***	
			[0.372]	
Intensity=2				2.589***
				[0.822]
Intensity=3				4.071***
				[0.872]
Intensity=4				5.054***
				[1.386]
Intensity=5				7.386***
				[2.202]
Specialty FE	N	Y	N	N
Cluster FE	N	N	Y	Y
Adjusted R^2	0.002	0.097	0.942	0.940
Observations	4181	4181	78	78
Statistics of totRVU and wRVU/totRVU				
	Sample	Mean	Median	St. Dev.
totRVU	All	16.21	9.5	22.93
	E/M	3.09	2.88	1.44
wRVU/totRVU (%)	All	41.27	44.55	20.26
	E/M	66.16	66.35	10.91

Table A25: Variation in (wRVU/totRVU) across code intensities

NOTES: We restrict the sample to HCPCS codes with nonzero total RVUs (totRVU). All RVU values are taken from the 2012 Physician Fee Schedule using non-facility pricing amounts. The dependent variable throughout is wRVU/totRVU measured in percents (%). Columns (1) and (2) show how wRVU/totRVU changes with totRVU in general, without and with code specialty fixed-effects, respectively. Columns (3) and (4) focus on E/M codes and controls for code cluster (consisting of 3-5 codes) fixed-effects. Column (3) assumes a linear effect of code intensity on wRVU/totRVU, whereas Column (4) includes a dummy variable for each level of intensity except for the baseline (Intensity=1). Heteroskedasticity robust standard errors of the mean estimator are reported in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The bottom panel shows summary statistics of totRVU and wRVU/totRVU that are helpful in interpreting the results. The "Sample" columns specifies whether the statistics are calculated on the sample of 4,181 HCPCS codes or on the subsample of 78 E/M codes.

	NAMCS	CMS
1(MD)	0.941	0.940
	(0.005)	(000)
1(solo practice)	0.367	0.307
	(0.012)	(0.001)
1(in IT incentive program)	0.423	0.543
	(0.012)	(0.001)
No. of Unique Physicians	3,583	472,110

Table A26: NAMCS and CMS physician characteristics
NOTES: All NAMCS-related statistics are weighted. Standard errors of the mean estimator are reported in parentheses.

These effects are qualitatively similar to those found in Column (3) and are also fairly small.

In conclusion, wRVU/totRVU does not vary significantly with code intensity or total RVU. Hence we are confident that the lower reported hourly revenues of flagged physicians relative to their unflagged peers are not a mechanical result of RVU composition.

I Comparability of the Main Sample and the NAMCS Sample

Table A26 shows the balancing test results between the NAMCS 2012 data and our main sample constructed from the CMS Medicare Part B FFS Physician Utilization and Payment Data. A few things are done to ensure the comparison between the two is sensible. NAMCS restricts its sample to Doctors of Medicine (MD) and Doctors of Osteopathy (DO). The CMS sample is thus also restricted to include only those with an MD or DO. A tiny fraction of physicians are both MD and DO (59 in total) and they are counted as DOs for calculations in this table. "Solo practice" in NAMCS questionnaires is not explicitly defined. Thus 1(solo practice) in the CMS sample is defined as having no more than 5 providers (including nurses and physician assistants, etc.). NAMCS only asks the sampled physician whether the (group) practice they belong to has *applied* for CMS incentive programs encouraging effective use of health IT. Thus 1(in IT incentive program) in CMS is defined accordingly as a dummy variable for participation in *any* of the incentive programs.

References

American Academy of Physician Assistants, "Reimbursement Issues: Thrid-Party Reimbursement for PAs," 2010.

Centers for Medicare and Medicaid Services, "Medicare Carriers Manual Part 3 - Claims Process, Transmittal 1764," August 2002.

Department of Health and Human Services, "HIPPA Administrative Simplification: Standard Unique Health Identifier for Health Care Providers; Final Rule," 2004.

Gabbert, W, KH Kachur, and T Whitehead, *Current Procedural Coding Expert*, Salt Lake City, UT: Ingenix, 2012.

Gustafson, Shanna L, Gail Pfeiffer, and Charis Eng, "A Large Health System's Approach to Utilization of the Genetic Counselor CPT® 96040 code," *Genetics in Medicine*, December 2011, *13* (12), 1011–1014.

Pham, Hoangmai H., Deborah Schrag, Ann S. O'Malley, Beny Wu, and Peter B. Bach, "Care Patterns in Medicare and Their Implications for Pay for Performance," *New England Journal of Medicine*, 2007, *356* (11), 1130–1139. PMID: 17360991.

Welch, W Pete, Sally C Stearns, Alison E Cuellar, and Andrew B Bindman, "Use of Hospitalists by Medicare Beneficiaries: A National Picture," *Medicare & Medicaid Research Review*, 2014, *4* (2).

Wolman, Dianne Miller, Michael ME Johns, Cheryl Ulmer et al., *Resident Duty Hours: Enhancing Sleep, Supervision, and Safety*, National Academies Press, 2009.

Zuckerman, Stephen, Robert Berenson, Katie Merrell, Tyler Oberlander, Nancy McCall, Rebecca Lewis, Sue Mitchell, and Madhu Shrestha, "Development of a Model for the Valuation of Work Relative Value Units: Objective Service Time Task Status Report," Centers for Medicare & Medicaid Services, 2014.