

Detecting Potential Overbilling in Medicare Reimbursement via Hours Worked[†]

By HANMING FANG AND QING GONG*

We propose a novel and easy-to-implement approach to detect potential overbilling based on the hours worked implied by the service codes which physicians submit to Medicare. Using the Medicare Part B Fee-for-Service (FFS) Physician Utilization and Payment Data in 2012 and 2013 released by the Centers for Medicare and Medicaid Services, we construct estimates for physicians' hours spent on Medicare beneficiaries. We find that about 2,300 physicians, representing about 3 percent of those with 20 or more hours of Medicare Part B FFS services, have billed Medicare over 100 hours per week. We consider these implausibly long hours. (JEL H51, I13, I18, J22)

Medicare benefit payments in 2014 totaled \$597 billion, accounting for 14 percent of the United States federal budget forecasts (Kaiser Family Foundation 2015; Congressional Budget Office 2014): as more baby boomers retire, government health care program expenditures will further increase and will account for 14 percent of the US gross domestic product (GDP) by 2039. From a public policy perspective, mitigating the inefficiencies in the Medicare system to ensure that every Medicare dollar is put to the best use is of first-order importance. This paper is about a particular form of inefficiency that is broadly referred to as *overbilling*, where providers file improper claims in order to increase the reimbursement from Medicare or other insurance companies. The Office of Inspector General at the US Department of Health and Human Services loosely defines two common types of overbilling (formally referred to as *improper claims*): *upcoding* refers to billing codes reflecting a more severe illness than actually existed or a more expensive treatment than was provided; *overcharging* refers to charging for more units of a service than was provided, or charging for services not provided at all (see Department of Health and

* Fang: Department of Economics and Leonard Davis Institute of Health Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104, and NBER (e-mail: hanming.fang@econ.upenn.edu); Gong: Department of Economics and Leonard Davis Institute of Health Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104 (e-mail: qinggong@sas.upenn.edu). We are grateful to the two anonymous referees, Anupam Jena, Matthew Kahn, Alex Li, Dan Zeltzer, and participants at the 6th Biennial Conference of the American Society of Health Economists (ASHEcon 2016) and those at a seminar at the University of Melbourne for helpful comments and suggestions. Fang gratefully acknowledges the generous financial support from NSF grant SES-1122902. All remaining errors are our own. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

[†] Go to <https://doi.org/10.1257/aer.20160349> to visit the article page for additional materials and author disclosure statement(s).

Human Services 2015).¹ Lorence and Spink (2002) estimate that overbilling cost the Federal government about \$12 billion annually in the 1990s, and researchers have continued to find evidence of overbilling since then (e.g., Brunt 2011).

Efficient and cost-effective detection of overbilling, preferably at the individual *provider* level, is crucial to reduce overbilling. However, this remains a challenging task. Most papers in the literature measure overbilling by the differential probability that higher-level codes are billed relative to lower-level codes, or by the high percentile rank in the distribution of total Medicare reimbursement received by providers. But such measures could be confounded by factors such as selection on patient and provider characteristics. Medicare claims data, available in more recent years, enabled researchers to control for some, but not all, patient and provider heterogeneities. Rosenberg, Fryback, and Katz (2000) developed a Bayesian model to adaptively detect questionable claims using previous hospital claims that insurers already audited. However, new costly audits are required to apply the methods to new claims data. The Comprehensive Error Rate Testing (CERT) program by the Centers for Medicare and Medicaid Services (CMS) faces a similar challenge, because the program needs to hire experts to review a large sample of claims every year (see Centers for Medicare and Medicaid Services 2015a). Geruso and Layton (2015) identified upcoding at the market level using risk scores and variations in financial incentives for physicians.

In this paper, we propose a novel approach to efficiently detect, or at least flag, potential Medicare overbilling using conservative estimates of the *hours worked* implied by service codes which providers submit to Medicare for reimbursement. Our idea is very simple. Every provider has a *fixed* number of hours in any given period, and most of the service codes that are submitted for reimbursement require that the provider spends a certain amount of time with the patient. If the hours worked implied from the service codes which a provider submits to Medicare are implausibly long, the provider is suspicious for overbilling. Our approach to flag potential Medicare overbilling has several key advantages. First, the existing physician-level billing data are sufficient to implement this approach, thus no additional data collection is needed. Second, by focusing on the implied hours worked within a given time period, our approach is immune to confounding factors such as selection on patient conditions. Third, our approach is flexible in the sense that it can be automated, and can be easily extended to a more general setting with augmented data, for example, by including other components of Medicare and/or additional physician billing information, say, billings for beneficiaries of other insurance programs. We should also note that our calculation of implied physician hours worked is deliberately conservative for the moment, and it is certainly not fail-proof especially given some well-noted data limitations (see, for example, O’Gara 2014; Jones, Craft, and Fritz 2015). Nonetheless, we believe it can serve as a useful first step for effective and more targeted auditing to reduce Medicare overbilling.

We apply our approach to detect potential Medicare overbilling using two waves of Medicare Part B fee-for-service (FFS) physician payment data. We construct

¹ There is a third type of improper claims that bill for services which lack medical necessity, sometimes known as *utilization abuse*. Detecting utilization abuse could be much harder and potentially controversial, so the approach we propose in this paper only targets detecting upcoding and overcharging.

conservative estimates for physicians' implied hours worked treating Medicare Part B FFS beneficiaries in 2012 and 2013.^{2,3} We find that about 2,300 physicians in our sample billed for more than 100 hours per week for Medicare Part B FFS patients alone. We consider such long hours of work highly implausible and refer to these physicians as *flagged physicians*. A comparison with the unflagged physicians shows that flagged physicians are more likely to work in smaller group practices, more likely to be a specialist rather than a primary care physician, and provide both more and higher-intensity services. Results from simple regression analysis also suggest that the coding patterns of the flagged physicians are sensitive to variations in the marginal revenue of choosing a higher-intensity code. Interestingly, the revenues from these higher-intensity services are not enough to offset the *longer* implied hours needed to furnish them, resulting in substantially lower reported hourly revenues for the flagged physicians than for the unflagged physicians. Moreover, this large gap in hourly revenues is hard to reconcile using observable physician characteristics and geographical variations.

Our research is related to the literature on the prevalence and consequences of overbilling. Lorence and Spink (2002) surveyed organizational providers and found significant coding optimization, despite serious penalties if the fraudulent billing practices were found out. Angeles and Park (2009) showed that upcoding imposed unnecessary cost to the already expensive Medicare program, was especially serious for Medicare Advantage, and this problem may worsen as the 76 million baby boomers age. The Department of Health and Human Services, alarmed at the rapid increase in Medicare spending from 2001 to 2010, conducted an in-depth study on the coding trends of evaluation and management (E/M) services in 2012, and found some physicians consistently billed higher-level codes.

Our study is also related to the literature on the possible determinants of overbilling. Adams, Norman, and Burroughs (2002) noted that the long documented difficulty of billing may lead to more erroneous coding, which can provide room for fraudulent coding at the same time. Other factors examined include pressure from the management teams (Lorence and Spink 2002; Dafny and Dranove 2009); hospital ownership (Silverman and Skinner 2004); anti-fraud enforcement effort (Becker, Kessler, and McClellan 2005; Bastani, Goh, and Bayati 2016); fee differentials across codes (Brunt 2011; Bowblis and Brunt 2014); and information technology such as electronic health records (EHR) (Adler-Milstein and Jha 2014).

Finally, our paper is related to the recent growing list of papers that used the newly released CMS Physician Utilization and Payment data. Most studies that use this dataset look at utilization and/or payment patterns of a particular specialty or procedure (Bergman, Saigal, and Litwin 2015; Harewood, Foley, and Farnes 2014; Clair and Goyal 2015; Dusetzina, Basch, and Keating 2015; Ip et al. 2015; Ko et al. 2015; Lapps et al. 2016; Menger et al. 2015; Schmajuk, Bozic, and Yazdany 2014; Sutphin et al. 2014; Skolarus et al. 2015; Skolasky and Riley 2015). For example, Bergman, Saigal, and Litwin (2015) studied physician payments in general and

²CMS released two waves of the data to the public in April 2014 and May 2015, respectively.

³Details on how we construct the estimates for physicians' hours worked based on the service codes submitted to Medicare are described in Section II.

found that high physician earnings were mainly driven by more services furnished per patient instead of more patients.

The remainder of the paper is structured as follows. In Section I, we describe the data and the construction of our sample. In Section II, we discuss our approach to estimating physician hours worked. In Section III, we present our empirical results. In Section IV, we corroborate some of our findings using two external datasets, the National Ambulatory Medical Care Survey and the CMS Comprehensive Error Rate Testing (CERT) results. Finally, we conclude in Section V.

I. Data and Sample Construction

Our main data source is the Medicare Part B FFS Physician Utilization and Payment data released annually to the public by the CMS since April 2014. The two waves of data available now are derived from all Medicare Part B FFS claims made in 2012 and 2013, respectively. Each wave of data has about nine million records at the provider-place-service level. Providers are uniquely identified by their national provider identifier (NPI) and characterized by a limited set of basic information (e.g., address, individual or organization indicator, gender, and specialty). Places are categorized into *office* settings and *facility* (such as hospitals) settings, and reflect where the provider furnished a service. Services are identified by a five-digit alphanumeric code specified in the Healthcare Common Procedure Coding System (HCPCS).⁴ Hence, each observation is a summary of a provider's rendition of a service at a place within the calendar year, as well as the payment the provider received for these services.

We supplement the Physician Utilization and Payment dataset with three other publicly available datasets. First, we use the CMS Physician Compare database to get more detailed physician characteristics such as education background and group practice affiliations. Both datasets identify physicians by their NPI, which facilitates the matching process. Then we use the National Physician Fee Schedule to get the Relative Value Units (RVUs) that quantify the amount of work required to furnish each service, which we will use to estimate the time needed for the services. Finally, we use a CMS on-site survey (Zuckerman et al. 2014) that objectively measured the time needed for a subset of services to corroborate our estimates of physician hours worked.

Next, we construct our sample by including *only* providers who are individuals (physicians) instead of organizations, are matched with a record in the Physician Compare database, work in the continental United States, completed professional medical training between 1946 and 2011, and have valid basic information (practice location, gender, and specialty).⁵ The first selection criterion discards about 5 percent of observations in the raw Physician Utilization and Payment data, the

⁴For every physician, the HCPCS codes claimed for fewer than ten times in a calendar year are excluded from the datasets to protect patient privacy. Claims for durable medical equipment are also excluded.

⁵It is important in our analysis that all claims under the same NPI are services furnished by the same individual. The NPI Final Rule (Department of Health and Human Services 2004) ensures that this *should be* the case for individual NPIs. To the extent that some providers submit claims using another provider's NPI, we also consider it to be a form of inappropriate billing. See the online Appendix for more detailed institutional background regarding billing and individual provider NPIs.

second discards about 9 percent of the remaining observations, and the rest lead to only negligible reduction of the sample size.^{6,7} We further restrict our sample to include only HCPCS codes that are actually services; that is, we exclude codes that are drugs, equipment, or medical supplies, or are only for quality administration purposes and not paid for, or are temporary codes for new services.^{8,9}

Finally, we aggregate the data to the physician-service level. For each physician-service combination, we observe the physician's characteristics, the workload to furnish the service (RVU and/or time needed), the volume of the service billed each year (the number of times that the service is furnished and the number of unique Medicare Part B FFS beneficiaries receiving the service), and total Medicare payments for these services. In the final sample, we have 7.9 million physician-service observations on 623,959 physicians and 4,480 HCPCS service codes.

II. Measuring Physician Hours Worked

We define *physician hours worked* to be the total time a physician spent with patients to furnish the service codes submitted to Medicare for reimbursement. For some codes, referred to below as *timed codes*, the time needed per service is readily available; for other service codes that do not have such information, we use the timed codes to estimate the time needed for providing these services based on their Work Relative Value Units (Work RVUs) which are measures of workloads that CMS assigns to all services.

A. Timed Codes

The timed codes, which are the cornerstone of our estimation of physician hours worked, fall into two categories. The first category is 145 timed codes with suggested or required time in their definition, and accounts for 50.44 percent of total Medicare Part B FFS reimbursement in 2012. These are mostly from the evaluation and management (E/M) code group, which include office or home visits. An important feature of these services is that there are usually multiple codes with different levels of intensity or complexity to furnish even for a narrowly defined service, and the physician has discretion over which one to file. The American Medical Association (AMA) publishes guidelines on choosing the most appropriate code, and usually includes typical time needed for E/M codes (Gabbert, Kachur, and Whitehead 2012). As an example, Table 1 shows a typical cluster of E/M codes where multiple codes are available for the *same* service but have varying workload requirement and fees. All five HCPCS codes, 99201 through 99205, are for "office or other outpatient

⁶For expositional simplicity, we will refer to all individual providers as *physicians* even if a small fraction of them are nurses or physician assistants.

⁷We exclude those graduated in or after 2012 because they are likely to be residents, who are known for extremely long working hours. We discuss this in greater detail in the online Appendix and show that our results are not affected when more possible residents are excluded from the sample.

⁸Drug codes are excluded when we estimate total physician hours worked but not when we calculate total revenues.

⁹We exclude temporary codes because they do not have RVU information that we can use to reliably estimate the time needed to furnish them.

TABLE 1—EXAMPLE OF CODES WITH VARYING INTENSITY AND TIME NEEDED FOR THE SAME SERVICE

HCPCS	Coding requirement			Typical time needed	Work RVU	2012 price (\$)
	History	Exam	MDM			
99201	PF	PF	Straightforward	10 minutes	0.48	31.09
99202	EPF	EPF	Straightforward	20 minutes	0.93	53.54
99203	Detailed	Detailed	Low	30 minutes	1.42	77.47
99204	Comprehensive	Comprehensive	Moderate	45 minutes	2.43	118.18
99205	Comprehensive	Comprehensive	High	60 minutes	3.17	145.81

Notes: All five codes are for office or other outpatient visits for new patients. The 2012 prices are for services furnished in office settings prior to the adjustment using geographic practice cost indices (GPCI). History refers to the process of asking about a patient's health history; Exam refers to the physical examinations; MDM refers to the complexity of medical decision making. To file a certain code, all three elements of the service must meet the required level. PF (problem-focused) is limited to the affected area or organ system; EPF (expanded problem-focused) also includes related areas. CMS provides very detailed definitions of each level of service intensities in (Centers for Medicare and Medicaid Services 2015b), to which we refer the interested reader.

visits for new patients.” But the lowest intensity code, 99201, only needs 10 minutes to furnish per the AMA guidelines, and generates \$31.09 of revenue, whereas the highest intensity code, 99205, needs 60 minutes and generates \$145.81.¹⁰ Note that, incentive issues aside, if a physician were to overstate the service intensity by one level, revenue would increase by at least \$20.

The second category is codes selected in a 2014 CMS survey that directly measures the time needed for certain services (Zuckerman et al. 2014). The survey targets 112 HCPCS codes that are judged to be growing fast, frequently billed, or often billed together.¹¹ These codes make up 17.76 percent of 2012 Medicare Part B FFS reimbursement. Survey staff are sent on site to document the time used to furnish the interested services at several participating institutions with large volumes of these services.¹²

Our idea is to use the time needed for timed codes described above to estimate the time needed for all other codes. In order to do this, we construct the *expected time needed* for each code based on the typical time needed suggested by the AMA guideline.¹³ This is important because the actual time to furnish a service code may vary both across and within physicians. We construct the expected time needed as follows. Assuming the time needed follows a uniform distribution, we take the simple average of the minimum and maximum time allowed for each code to get the expected time. Specifically, some codes may have an explicit range of time needed, such as “5–10 minutes of medical discussion.” For such codes, the expected time needed is simply the average of the lower and upper bounds. For codes that do not have such a range, physicians are supposed to file the code whose typical time needed is closest to the actual time spent. For example, between codes 99202 and

¹⁰These fees are the baseline reimbursement amounts in the 2012 Physician Fee Schedule. Actual Medicare payments will vary slightly across geographic regions and specific settings in which the services are furnished.

¹¹One of the 112 HCPCS codes in the 2014 CMS Survey is a Level-II code (mainly products and supplies) with a service component, which we exclude when constructing our main sample.

¹²The representativeness of specialties among the selected codes is discussed in the online Appendix.

¹³An exception is when the AMA guideline *requires* the physician to spend a certain amount of time when furnishing a service. For example, code 99360 is for “physician standby service, requiring prolonged physician attendance, each 30 minutes (e.g., operation standby, standby for frozen section, for cesarean/high-risk delivery, for monitoring EEG),” and explicitly prohibits filing this code for services less than 30 minutes.

99203 as described in Table 1, a physician who spent 23 minutes should file the code 99202 instead of 99203. Following this logic, the expected time needed we will assign to HCPCS codes 99201 through 99205 are 7.5, 20, 31.25, 45, and 60 minutes, respectively. To see this, consider HCPCS code 99201 for example. Note that physicians who spend 0 to 15 minutes with a new patient are supposed to file HCPCS code 99201 if they follow the AMA guideline. Thus, under the plausible assumption that the actual time spent with patients follows a uniform distribution, the simple average of the minimum (0 minutes) and maximum (15 minutes) time allowed for filing 99201 is 7.5 minutes. For the highest intensity codes within a cluster of codes, e.g., code 99205 in Table 1, we set the expected time to be the same as the typical time as there is no upper bound specified in the AMA guideline. In order to err on the conservative side, we moreover choose the *smaller* of the typical time for a service code and expected time we construct for the timed codes whenever the two differ. Finally, we also exclude timed codes that do not require direct contact of the physician with the patient, such as intravenous drug infusions and phototherapies, again to be conservative in our estimate of the physicians' hours worked.

B. Work RVUs and Time Needed for Untimed Codes

Next, we estimate the time needed for all other codes for which the AMA guideline does not specify typical or required time. Our estimation is based on the relative value units (RVUs) of the service codes. RVUs reflect the value of each HCPCS code along three dimensions. AMA appoints a special committee of experts from various specialties to assign and regularly update the RVUs; and the CMS uses them to determine Medicare reimbursements to physicians. The Physician Fee Schedule specifies the following formula for the baseline payment amount for a given HCPCS code:

$$\text{Payment} = [(\text{Work RVU}) \times (\text{Work GPCI}) + (\text{PE RVU}) \times (\text{PE GPCI}) \\ + (\text{MP RVU}) \times (\text{MP GPCI})] \times \text{CF},$$

where Work RVU captures the amount of work, primarily time, needed to furnish the service; PE RVU captures the practice expense (PE) of the service; MP RVU captures the malpractice insurance cost of the service; the geographic practice cost indices (GPCI) adjust for the geographic differences in the costs of practicing medicine, and vary slightly around 1 across 90 GPCI regions in the US; and finally, the conversion factor (CF) translates the RVUs into dollar amounts, which is \$24.6712 per RVU in 2012, and \$34.023 per RVU in 2013 (Centers for Medicare and Medicaid Services 2013).

We use Work RVUs to estimate the time needed for untimed codes. Though Work RVUs are imperfect measures of service time, time still plays the central role in how Work RVUs are determined, making Work RVUs the best available tool for our purposes.¹⁴ Our estimation takes two steps. First, we take the timed codes, for which we know the time needed and the Work RVUs, and estimate the time needed

¹⁴ See Centers for Medicare and Medicaid Services (2014). Merrell et al. (2016) also show that time explains 80–90 percent of variation in work RVUs.

per Work RVU. We use both the simple averages and regressions for robustness, and allow for both the intercept and the coefficient on work RVU to differ by the 15 service code groups that a HCPCS code can belong to in order to account for the potential difference in the practice patterns across specialties.¹⁵ Second, we use the Work RVU of each code, whether it is a timed code or an untimed code, and the estimated time needed per Work RVU from the first step to calculate its time needed.

Thus, for every timed code, we will have three measures of time needed: its expected time needed according to the AMA guideline, and two estimated time needed using the two estimation methods in the first step; and for every untimed code, we will have the latter two measures of time needed. Again to err on the conservative side, we pick the minimum of the measures of the time needed for each service code.¹⁶ In the end, we get positive estimates of time needed for 75 percent of the HCPCS codes. Codes that do not have positive estimates of time needed, which we will refer to as *zero-time codes*, are either drugs/supplies that do not require direct contact with the physician (thus have zero Work RVU), or those that have negative estimates of time needed according to our procedure.

Given the time needed estimates of the HCPCS codes, we calculate physician i 's total hours worked in year t based on services i billed to Medicare in calendar year t :

$$(1) \quad (\text{Hours worked})_{it} = \sum_{j \in J} [(\text{Time needed per service } j) \times (\text{Number of service } j \text{ billed})_{it}],$$

where j is a HCPCS code in the set of codes, J , for which we have obtained positive time needed using the procedure described above.

C. Discussion of Estimated Physician Hours Worked

Our estimates of physician hours worked are likely to be a *lower bound* of the actual hours, provided that the service codes truthfully reflect both the volume and the intensity of services the physicians actually furnished. We acknowledge the possibility of imprecise estimates, especially with the relatively small set of timed codes that we could use. That said, we would like to note that the timed codes already account for over two-thirds of total Medicare Part B FFS reimbursement, leaving limited room for the imputed time to affect the key findings; moreover, we have taken a series of measures to prevent overflagging of potential overbilling. First, as we described above we make every decision in the construction of the hours worked to err on the conservative side. Second, the Medicare Part B FFS Physician Utilization and Payment data that we use in our estimation only include Medicare Part B FFS claims, which on average account for less than 31 percent of a physician's services (see Physicians Foundation 2012). Third, as we mentioned in

¹⁵We also include the code group fixed effects to control for possible distortions in the assignment of RVUs across specialties. Some studies (e.g., Chan and Dickstein 2016) found that better-represented specialties get more favorable assignment by the RVU Update Committee.

¹⁶In particular, the service time of *timed codes* used in our estimation of physician hours worked are on average 88.44 percent of the typical time for the timed codes from the AMA Guidelines, and are on average 93.53 percent of the measured time for the timed codes from the 2014 CMS survey.

footnote 4, to protect patients' privacy the dataset released by CMS excludes, for each physician, the HCPCS codes that were claimed by the physician for fewer than ten times in a calendar year. Fourth, we only include the time needed for 75 percent of HCPCS codes that represent services requiring physicians' direct contact with the patient, have nonzero work RVUs, and end up with positive time estimates. Finally, sometimes physicians bill under the NPI of organizational providers (e.g., a hospital or a group practice), which we exclude from our analysis because it is impossible to identify an individual physician's contribution to the organizations' billing records.¹⁷

III. Describing Physician Hours Worked

We convert the total hours worked in year t to hours worked per week in year t for easier interpretation in the analysis that follows, assuming physicians work 51 weeks each year (i.e., take only 9 days off). By doing so we essentially characterize physician hours worked per week *averaged* over the entire year. Hence, we are allowing the physicians to possibly smooth hours of work intertemporally during the year, which of course is another conservative choice that is likely to lead to underdetection of overbilling. This choice is necessitated by the data limitation that utilization and payment records are aggregated to the calendar year level, and not at a higher time frequency. However, if we are able to detect implausibly long hours worked per week under the lenient criterion permitting smoothing across the weeks, it would serve as a stronger signal for potentially inappropriate coding behavior.

Figure 1 graphs the distribution of the average reported hours worked per week across all the physicians. Despite the conservative methods we used to estimate the physicians' hours worked, about 2,300 physicians submitted claims for service codes that would translate into over 100 hours per week on services for Medicare Part B FFS beneficiaries. Moreover, about 600 physicians submitted claims for service codes that would imply over 168 hours per week (i.e., 24 hours per day and 7 days per week). To put these numbers into perspective, the Accreditation Council for Graduate Medical Education (ACGME) restricts residency working hours to 80 hours per week since 2003 in light of the much-studied sleep deprivation and performance deterioration of health care providers (Institute of Medicine of the National Academies 2009).

In Table 2 we take a closer look at physicians who billed Medicare for long hours. We use different *flagging* thresholds in terms of weekly hours, 80, 100, 112 (16 hours per day for 7 days), and 168 (24 hours per day for 7 days), respectively, and present the statistics by year. We will refer to physicians whose estimated weekly hours worked above the threshold as *flagged physicians* and those below as *unflagged physicians*. For example, under the 100 weekly hours threshold, we flag 2,292 physicians in 2012 and 2,120 physicians in 2013 as having submitted claims with implied hours worked exceeding that threshold. They account for 2.71 percent and 2.55 percent of all physicians in our data who have submitted claims implying

¹⁷In addition, we show in the online Appendix that all key findings remain under four alternative methods of estimating hours worked, including using the *minimum* required time for timed codes only.

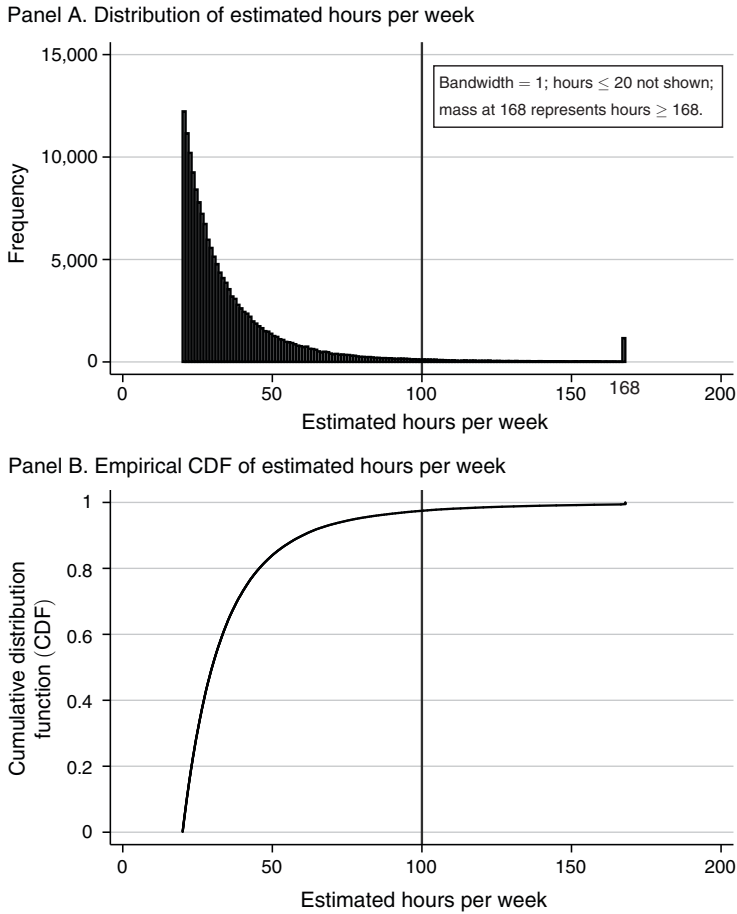


FIGURE 1. DISTRIBUTION OF ESTIMATED HOURS PER WEEK

Notes: The horizontal axis shows physicians’ hours worked per week estimated from the claims they submitted to Medicare for reimbursement. We restrict the sample to physicians whose claims imply at least 20 hours per week in at least one year. The mass at 168 per week in the top panel represents hours larger than or equal to 168 per week.

at least 20 hours of service per week in at least one year, and 0.367 percent and 0.340 percent among all physicians in 2012 and 2013, respectively.

Physicians with very few implied hours worked in our sample could have few Medicare patients, or could have just as many Medicare patients but they specialize in the 25 percent zero-time service codes where information on the time needed is unavailable. If it is the latter, one might be concerned that our results overlook physicians who are only overbilling on the zero-time codes. Table 2 indeed shows that, for example, the number of distinct zero-time HCPCS codes as a fraction of all 4,480 distinct HCPCS codes filed by flagged physicians ranges from about 9 percent using the 80-hour threshold to about 4 percent using the 168-hour threshold, while the corresponding fraction for unflagged physicians is over 13 percent using any threshold. However, when weighted by Work RVUs of the service codes, the differences are much smaller. This is consistent with the fact that many of the zero-time codes have low Work RVUs. Similarly, the differences between flagged

TABLE 2—NUMBER AND FRACTION OF PHYSICIANS FLAGGED

Hours threshold:	80+		100+		112+		168+	
	2012	2013	2012	2013	2012	2013	2012	2013
Number of physicians flagged	4,125	3,838	2,292	2,120	1,689	1,546	615	530
Fraction in physicians working 20+ hr/week	4.879	4.618	2.711	2.551	1.998	1.860	0.727	0.63
Fraction in all physicians	0.661	0.615	0.367	0.340	0.271	0.248	0.099	0.085
Fraction of zero-time codes (flagged)	9.008	9.147	7.352	7.377	6.430	6.683	4.141	4.250
Fraction of zero-time codes (unflagged ^a)	13.17	13.30	13.16	13.30	13.15	13.30	13.14	13.30
Fraction of wRVU from zero-time codes (flagged)	0.171	0.172	0.115	0.108	0.107	0.098	0.033	0.054
Fraction of wRVU from zero-time codes (unflagged ^a)	0.591	0.611	0.590	0.611	0.590	0.611	0.590	0.611
Fraction of volume from zero-time codes (flagged)	11.43	11.85	9.173	9.391	8.152	8.134	4.104	4.421
Fraction of volume from zero-time codes (unflagged ^a)	15.70	15.83	15.69	15.83	15.68	15.83	15.67	15.83
Fraction of revenue from zero-time codes (flagged)	8.112	9.306	7.106	8.025	6.372	7.018	3.557	3.772
Fraction of revenue from zero-time codes (unflagged ^a)	7.442	7.743	7.435	7.743	7.432	7.743	7.426	7.743
Total number of physicians working 20+ hr/week					96,033			
Total number of physicians					623,959			

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. Fraction in physicians working 20+ hr/week shows the fraction of flagged physicians among physicians who billed at least 20 hours per week in the same calendar year. Fraction in all physicians shows the fraction among all physicians in our sample, which covers the vast majority of physicians. Zero-time codes are codes for which positive time needed estimates are not available and account for 25 percent of all HCPCS codes. wRVU is the physician work RVUs that are specific to each HCPCS code and reflect the amount of work (primarily time) required to furnish each service.

^aUnless otherwise specified, unflagged refers to unflagged physicians whose estimated weekly hours worked are above 20 in 2012 or 2013 or both. All fractions are measured in percent.

TABLE 3—FLAG PATTERNS ACROSS TIME

Hours threshold:	80+		100+		112+		168+	
	Count	Share (percent)	Count	Share (percent)	Count	Share (percent)	Count	Share (percent)
<i>Years flagged</i>								
2012 only	1,135	27.52	704	30.72	539	31.91	233	37.89
2012 and 2013	2,990		1,588		1,150		382	
2013 only	848	22.09	532	25.09	396	25.61	148	27.93

Notes: Hours threshold shows the cutoff number of hours billed per week above which a provider is flagged. Count columns report the number of physicians flagged (in 2012 only, in both years, or in 2013 only). Share columns show the fraction of physicians who are only flagged in 2012 (2013) among all physicians flagged in that year (percent).

and unflagged physicians are big when we compare the unweighted number of claims for zero-timed codes, but shrink significantly once weighted by revenues. These suggest that specialization in different types of services is unlikely to result in large underflagging of physicians, though having differential fractions of Medicare patients still is.¹⁸

Table 3 decomposes the flagged physicians into those flagged in 2012 only, in 2013 only, and in both years. For example, using the 80-hour threshold, 1,135 physicians are flagged in 2012 only, accounting for 27.52 percent of the 4,125 physicians flagged in 2012; 848 physicians are flagged in 2013 only, representing 22.09 percent of the 3,838 physicians flagged in 2013.

¹⁸We can not address this issue given the fact that our data only contains Medicare claims. However, our method can be easily extended to a more general setting with augmented data from physician billing information for beneficiaries of other insurance programs.

A. Who Reported Implausibly Long Hours?

In Tables 2 and 3, we used four different weekly hours thresholds to flag physicians. From now on, we focus on physicians flagged using the 100-hour threshold, although all results hold under alternative flagging criteria.¹⁹ In addition, we only focus on the subsample of 96,033 physicians with more than 20 hours worked per week treating Medicare Part B FFS patients, as we believe this is the more relevant group to be compared with the flagged physicians for reasons discussed earlier.

In Table 4 we compare the characteristics of the following groups of physicians, according to their respective column headings: (1) flagged in any year; (2) flagged in 2012; (3) flagged in 2013; (4) flagged only in 2012; (5) flagged in both 2012 and 2013; (6) flagged only in 2013. Column 7 shows the means for the baseline group, those never flagged. To account for the heterogeneity in physicians' exposure to local Medicare markets, we compare the characteristics of physicians controlling for hospital referral region (HRR) fixed effects. The 306 HRRs represent local health care markets and are commonly used as the unit of analysis for regional variations of health care in the United States (Wennberg and Cooper 1996). The number in each cell of the table is the estimated coefficient from an ordinary least squares (OLS) regression on the subset of physicians who are either never flagged, or have the flag status indicated by the column heading. We use the physician characteristic in the corresponding row as the dependent variable, and the flag status dummy as the explanatory variable together with HRR fixed effects. For example, physicians ever flagged are 3.2 percentage points more likely to be male than never flagged physicians, 17.8 percentage points less likely to have an MD, and tend to practice in groups with 52.3 fewer providers, etc.

We find that flagged physicians are slightly more likely to be male, non-MD, more experienced, and provide fewer E/M services. Importantly, they work in substantially smaller group practices (if at all), and have fewer hospital affiliations.²⁰ These characteristics are similar to what Cutler et al. (2015, p. 3) found about physicians who "consistently and unambiguously recommended intensive care beyond those indicated by current clinical guidelines."²¹

B. What Are the Specialties of Flagged Physicians?

In addition to the individual characteristics of flagged physicians, we are also interested in whether some specialties are more likely to be associated with flagged physicians. For this purpose, we follow Fryer and Levitt's (2004) approach toward

¹⁹We present the results under the 112 and 168 weekly hours flagging thresholds in the online Appendix.

²⁰The CMS Physician Compare data, from which we obtain the physicians' characteristics, do not report their race and age. Also note that the coefficients measure correlational and not necessarily causal effects.

²¹The differences between flagged and unflagged physicians in the sizes of group practice and the number of hospital affiliations could be due to various factors such as internal control, auditing stringency, and patient selection. It may suggest a group effect where overbilling is *contagious* within a group (e.g., Gino, Ayal, and Arieli 2009) or even across groups (e.g., via endogenous referrals as in Zeltzer 2016). It is also possible that large group practices might have a more diverse pool of patients than small practices so that the share of Medicare Part B FFS patients represent a smaller fraction of their total patients for physicians working in larger practices. Unfortunately, with the current set of Medicare Part B FFS data we could not test which of the aforementioned factors are driving the observed differences between flagged and unflagged physicians.

TABLE 4—CHARACTERISTICS OF FLAGGED PHYSICIANS VERSUS UNFLAGGED PHYSICIANS, CONDITIONAL ON HOSPITAL REFERRAL REGION

	Ever (1)	2012 (2)	2013 (3)	2012 only (4)	Both (5)	2013 only (6)	Never (7)
1(male)	0.032 [0.007]	0.037 [0.007]	0.037 [0.008]	0.015 [0.011]	0.046 [0.009]	0.007 [0.018]	0.856
1(MD)	-0.178 [0.034]	-0.171 [0.037]	-0.126 [0.032]	-0.338 [0.053]	-0.100 [0.036]	-0.217 [0.038]	0.844
Experience (years)	0.452 [0.245]	0.799 [0.278]	0.046 [0.252]	1.624 [0.602]	0.429 [0.279]	-1.244 [0.593]	24.124
Number of providers in group	-52.349 [5.991]	-49.660 [6.139]	-52.027 [5.981]	-53.157 [10.174]	-48.119 [5.969]	-65.027 [10.346]	88.919
Number of hospital affiliations	-1.392 [0.124]	-1.376 [0.130]	-1.338 [0.114]	-1.565 [0.233]	-1.300 [0.116]	-1.476 [0.170]	2.813
1(in Medicare)	0.004 [0.007]	0.003 [0.008]	0.001 [0.008]	0.014 [0.012]	0.000 [0.009]	0.009 [0.016]	0.857
1(in ERX)	-0.003 [0.016]	0.007 [0.019]	0.030 [0.016]	-0.107 [0.026]	0.057 [0.020]	-0.061 [0.025]	0.463
1(in PQRS)	-0.001 [0.016]	0.002 [0.017]	0.026 [0.016]	-0.086 [0.025]	0.041 [0.017]	-0.022 [0.025]	0.396
1(in EHR)	-0.028 [0.014]	-0.023 [0.016]	-0.032 [0.015]	-0.012 [0.029]	-0.028 [0.017]	-0.045 [0.024]	0.417
Types of codes 2012	0.220 [0.905]	0.966 [1.027]	1.460 [0.889]	-3.611 [1.494]	2.906 [1.045]	-3.495 [0.934]	22.428
Types of codes 2013	0.393 [0.905]	0.872 [1.018]	2.059 [0.901]	-4.705 [1.340]	3.248 [1.047]	-2.034 [1.005]	22.351
Types of E/M codes 2012	-1.826 [0.193]	-1.814 [0.203]	-1.793 [0.199]	-1.945 [0.323]	-1.764 [0.212]	-1.900 [0.247]	6.228
Types of E/M codes 2013	-1.780 [0.193]	-1.792 [0.201]	-1.683 [0.198]	-2.087 [0.313]	-1.671 [0.210]	-1.737 [0.245]	6.207
Number of physicians in group	2,824	2,292	2,120	704	1,588	532	93,209

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 on the subset of physicians who are either never flagged, or have the flag status indicated by the column heading. We use the physician characteristic in the corresponding row as the dependent variable, and the flag status dummy as the explanatory variable together with HRR fixed effects. Physician experience is imputed from the year of graduation. Number of providers in group refers to the number of providers in the group practice where the billing physician works and 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. Standard errors clustered at the HRR level are in brackets.

quantifying the “blackness” of first names, and construct the Specialty Flag Index (SFI) for specialty s :

$$(2) \quad \text{SFI}_s = \frac{100 \times \Pr(s | \text{flagged})}{\Pr(s | \text{flagged}) + \Pr(s | \text{unflagged})},$$

where the conditional probability $\Pr(s | \text{flagged})$ is defined as the fraction of flagged physicians in specialty s among all flagged physicians, and $\Pr(s | \text{unflagged})$ is the

TABLE 5—PHYSICIAN SPECIALTIES AND FLAG STATUS

	Fraction in all	Number unflagged		Number flagged		SFI	
		2012	2013	2012	2013	2012	2013
Optometry	1.893	1,252	1,390	566	428	94.87	93.17
Dermatology	4.185	3,557	3,525	463	495	84.19	86.15
Ophthalmology	7.960	7,258	7,260	386	384	68.50	70.09
Pathology	2.746	2,567	2,578	71	60	53.08	50.76
Nephrology	4.900	4,607	4,615	99	91	46.78	46.62
Cardiology	11.120	10,543	10,579	136	100	34.54	29.51
Internal medicine	11.089	10,573	10,567	77	83	22.95	25.81
All physicians		93,741	93,913	2,292	2,120		

Notes: The table shows seven specialties with the highest SFI, defined in equation (2), among the specialties with at least 50 flagged physicians. Fraction in all shows the percentage 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 and unflagged physicians by year in our sample.

fraction of unflagged physicians in specialty s among all unflagged physicians.²² The index ranges from 0 to 100. If all physicians in specialty s are flagged, then SFI_s takes on a value of 100. If only unflagged physicians are specialty s , then SFI_s is 0. If flagged and unflagged physicians are equally likely to be in specialty s , then SFI_s is 50. If flagged physicians are four times as likely to be in specialty s than unflagged physicians, then $SFI_s = 100 \times 4/(4 + 1) = 80$. This measure is invariant to the fraction of the flagged physicians among all physicians, and to the overall popularity of the specialty among all physicians.

The SFI is a convenient summary of how a given specialty is represented among the flagged relative to its share in the entire physician population. A SFI of 50 indicates that specialty is fairly represented among the flagged, i.e., $\Pr(s | \text{flagged})$ and $\Pr(s | \text{unflagged})$ are both equal to the fraction of specialty s among all physicians. A SFI above 50 indicates that the specialty is overrepresented among the flagged physicians.

Table 5 ranks the top specialties with at least 50 flagged physicians by their SFIs. For example, optometry is considerably overrepresented among the flagged physicians, accounting for more than 20 percent of flagged physicians but less than 2 percent of all physicians, leading to SFIs over 90 in both years. On the contrary, internal medicine physicians are much underrepresented among the flagged physicians, with SFIs around 25. Moreover, note how the SFI as defined in (2) differs from the simple probability that physicians of a given specialty are flagged, namely $\Pr(\text{flagged} | s)$. For example, 386 (a mere 5 percent) of the 7,664 ophthalmologists are flagged in 2012, yet ophthalmology still gets a high SFI of 68.5 because it only makes up 7.96 percent of all physicians in our sample but contributes 16.8 percent to the 2,292 flagged physicians.

²²We use the self-reported primary specialty when a physician is in multiple specialties.

C. What Codes Do Flagged Physicians Tend to Bill?

Similarly to how we constructed specialty flag index, we can also construct the code flag index (CFI) for each HCPCS code j as follows:

$$(3) \quad \text{CFI}_j = \frac{100 \times \Pr(j | \text{flagged})}{\Pr(j | \text{flagged}) + \Pr(j | \text{unflagged})},$$

where the conditional probability $\Pr(j | \text{flagged})$ is defined as the number of claims for HCPCS code j filed by flagged physicians as a fraction of the total number of claims for all service codes filed by flagged physicians; and $\Pr(j | \text{unflagged})$ is the number of claims for HCPCS code j filed by unflagged physicians as a fraction of the total number of claims for all service codes filed by unflagged physicians. Like SFI, the CFI takes on values between 0 and 100; if a code is filed only by flagged physicians, then its CFI will be 100; and if a code is filed only by unflagged physicians, then its CFI is equal to 0. A code with a CFI of 50 indicates that it is filed by flagged and unflagged physicians at equal rates.

In Figure 2, we show that there is a nonlinear relationship between a code's CFI and the probability that it is filed by flagged physicians. For example, a HCPCS code with a 20 percent probability of being filed by flagged physicians can have a CFI over 75 if unflagged physicians file the code at a much lower rate than flagged physicians.

In Figure 3 we plot the distribution of CFIs among all HCPCS codes in our data.²³ Panel A is the unweighted distribution, which is roughly uniform. Panel B weighs the codes by their corresponding service volume, i.e., the total number of times they are filed. Panel C weighs the codes by their total Medicare reimbursement. A comparison between panels B and C shows that HCPCS codes with high CFIs do not necessarily have a lot of volumes, but they do have disproportionate costs to Medicare relative to their volumes.

In Figure 4 we compare the distribution of CFIs among flagged physicians and that of unflagged physicians, weighted by volume of service. By construction, flagged physicians do tend to report more high-CFI codes.

D. Decomposing the Long Hours and Quantifying Potential Overbilling

The long hours worked as implied by the flagged physicians' claims to Medicare can result from *high volumes* of services with a given distribution of service intensity (the *extensive* margin) and/or a larger fraction of *higher-intensity* services (the *intensive* margin). In Table 6, we examine the composition of hours billed by flagged and unflagged physicians.

Columns 1 and 2 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. Flagged physicians submit more than twice as many service claims to Medicare in a year as unflagged physicians,

²³Of the 4,480 HCPCS codes, about 1,800 have a CFI of 0 in either year, and about 220 have a CFI of 100. These codes are excluded from the figures so as not to distort the scales.

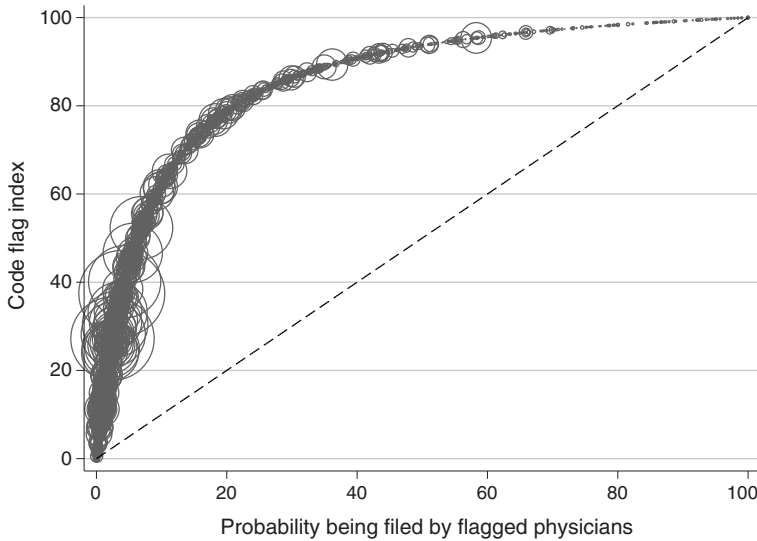


FIGURE 2. THE RELATIONSHIP BETWEEN HCPCS CODE FLAG INDEX AND ITS PROBABILITY OF BEING FILED BY FLAGGED PHYSICIANS

Notes: The horizontal axis shows the probability of the HCPCS code being filed by a flagged physician (in percent). 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. Each circle represents a HCPCS code, with the radius proportional to total revenue. The dashed line is the 45-degree line.

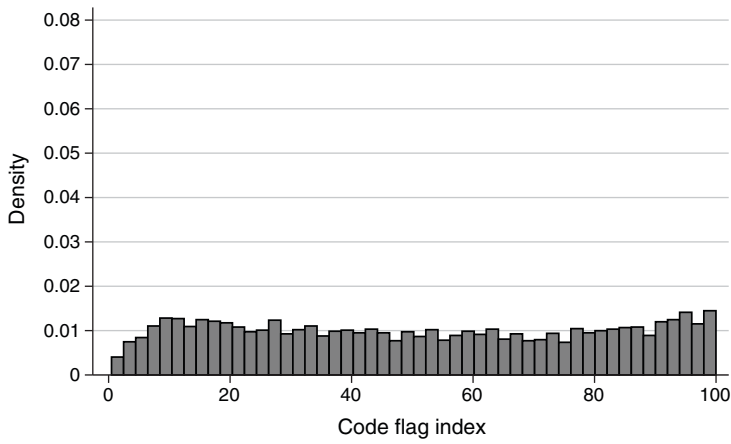
have about twice as many distinct Medicare patients in total, treat about twice as many Medicare patients per day, and provide significantly more services per patient. However, flagged physicians tend to file service claims with longer time requirement, resulting in fewer services furnished per hour than unflagged physicians. Interestingly, the implied Medicare payment per reported hour worked for flagged physicians is significantly lower than that for unflagged physicians. Taking 2012 for example, flagged physicians on average furnish 170 percent more services per year than unflagged physicians (7,708 + 4,540 = 12,248 versus 4,540); they have 106 percent more Medicare Part B FFS patients (2,571 + 2,430 = 5,001 versus 2,430); they provide 62 percent more services on each patient (1.67 + 2.43 = 4.10 versus 2.43). They also tend to provide higher-intensity services, which take longer to furnish (2.88 – 1.34 = 1.54 versus 2.88 services per hour, or 39 versus 21 minutes per service) and generate more revenue from Medicare payment (\$8.90 per service). However, since the higher revenue services require longer time to furnish, the Medicare payment per hour for flagged physicians are substantially (\$43.37 per hour) lower than their unflagged peers.

The sizable difference in Medicare payment per hour between flagged and unflagged physicians motivates our construction of the overbilling potential factor (OPF), which quantifies the extent to which there may be overbilling. We provide two alternative ways to construct the OPF.

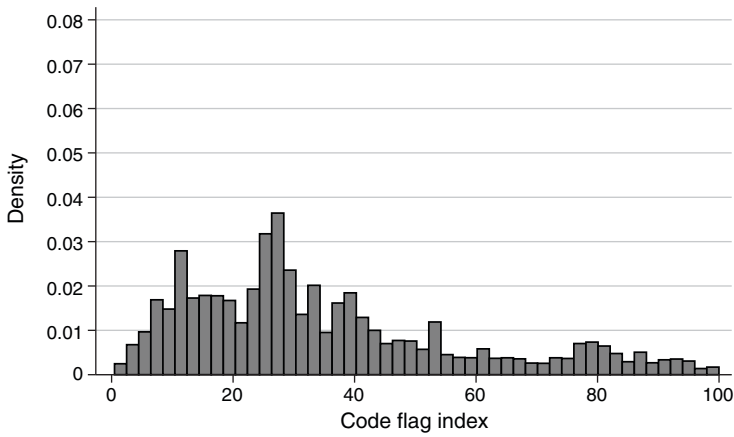
Our first measure of overbilling potential is

$$(4) \quad OPF1_i \equiv \frac{(\text{Total revenue})_i}{(\text{Fair revenue})_i} = \frac{(\text{Total revenue})_i}{(\text{Fair hourly revenue})_i \times (\text{Fair hours})}$$

Panel A. Unweighted



Panel B. Weighted by service volume



Panel C. Weighted by total Medicare reimbursement

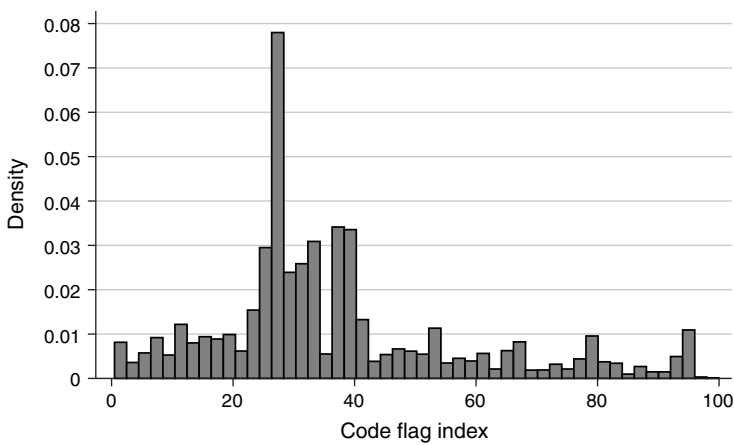


FIGURE 3. DISTRIBUTION OF HCPCS CODE FLAG INDEX

Notes: 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.

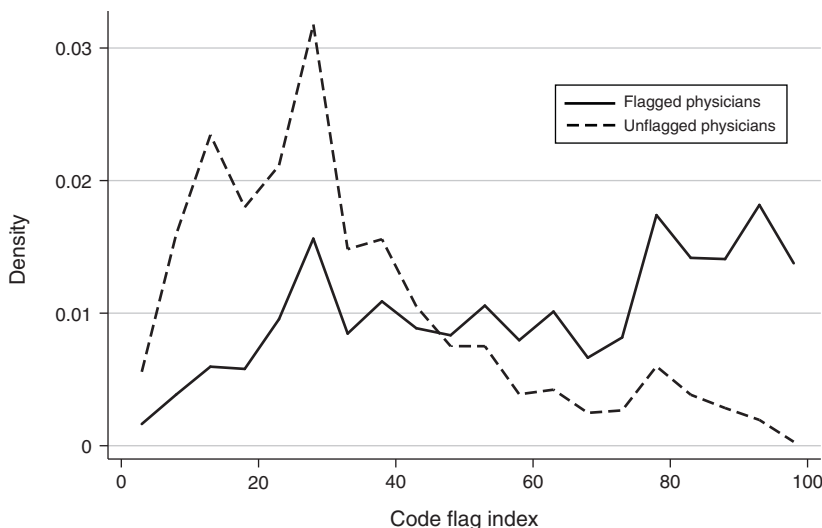


FIGURE 4. DISTRIBUTION OF CODE FLAG INDEX WEIGHTED BY SERVICE VOLUMES: FLAGGED PHYSICIANS VERSUS UNFLAGGED PHYSICIANS

Notes: 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.

where *Total revenue* is the observed annual Medicare Part B FFS payments of physician *i*; *Fair hourly revenue* is the predicted hourly revenue for physician *i* based on an OLS regression of the hourly revenues of *unflagged* physicians on observables, which include physician gender, credential, years of experience, and a full set of specialty, HRR, and year fixed effects; and *Fair hours* is set to be 8 hours per day multiplied by 365 days. An OPF1 above 1 captures the excess revenue relative to the predicted *fair* amount that is not explained by observed physician and local market characteristics.

Our second measure of overbilling potential is to compare the reported hours and the likely true hours, where the latter is the unknown number of hours physicians actually worked. Under the assumption that the goal of overbilling is to achieve the same revenue with fewer *actual* hours, we have, for each flagged physician *i*:

$$\begin{aligned}
 (\text{True Hours})_i \times (\text{Fair hourly revenue})_i &= (\text{Reported hours})_i \\
 &\times (\text{Reported hourly revenue})_i.
 \end{aligned}$$

Thus,

$$(5) \quad \text{OPF2}_i \equiv \frac{(\text{Reported hours})_i}{(\text{True hours})_i} = \frac{(\text{Fair hourly revenue})_i}{(\text{Reported hourly revenue})_i},$$

where, as in (4), *Fair hourly revenue* is the predicted hourly revenue for physician *i* based on an OLS regression of the hourly revenues of *unflagged* physicians on

TABLE 6—VOLUME OF SERVICES SUPPLIED CONDITIONAL ON HOSPITAL REFERRAL REGIONS:
FLAGGED VERSUS UNFLAGGED PHYSICIANS

	Flagged		Unflagged	
	2012	2013	2012	2013
Number of services provided	7,707.660 [801.128]	7,769.498 [785.011]	4,540.285	4,490.308
Number of services per patient	1.665 [0.253]	1.203 [0.191]	2.434	2.376
Number of services provided per hour	-1.340 [0.089]	-1.374 [0.081]	2.880	2.897
Number of patients	2,571.489 [452.642]	2,846.790 [463.427]	2,429.509	2,424.339
Number of patients per day	7.026 [1.237]	7.799 [1.270]	6.638	6.642
Number of patients per hour	-0.932 [0.057]	-0.933 [0.052]	1.577	1.591
Medicare payment per service (\$)	8.897 [3.758]	13.934 [3.346]	74.811	73.381
Medicare payment per patient (\$)	51.843 [8.750]	52.467 [6.642]	150.639	146.120
Medicare payment per hour (\$)	-43.367 [5.619]	-40.131 [4.808]	162.010	159.035
Observations	2,292	2,120	93,741	93,913

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. The last two columns report the means of the two unflagged groups as references. Number 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. Number of patients per day is the average number of patients per day assuming 366 (365, respectively) working days in year 2012 (2013, respectively). Per hour statistics are calculated using the estimated total hours worked of each physician.

observables, which include physician gender, credential, years of experience, and a full set of specialty, HRR, and year fixed effects; *Reported hourly revenue* is simply the total revenue received by physician i divided by the total hours reported by i , which we estimated based on i 's claims.

The OPFs corroborate the previous findings using the flag indicator by showing the discrepancy in revenues, total or hourly, that are unexplainable by physician observables.²⁴ Because the construction of the OPFs uses a scale-free measure (the hourly revenue), a key advantage of the OPFs is that they mitigate the problem that we do not know the shares of Medicare service in physicians' total workload. Therefore, it is possible that a physician is not flagged based on implied hours worked previously because Medicare is only a very small fraction of his/her patient pool, yet he/she may be flagged based on the OPFs if the physician's reported hourly revenue among his/her Medicare billing is too low relative to the fair hourly revenues.

²⁴Using the potentially untruthful *billed* hours is precisely motivated by this purpose. With a slight abuse of terminology, the *null hypothesis* that the flagged physicians are truthfully reporting hours is *rejected* if their implied hourly revenues are much lower than their peers with similar characteristics.

TABLE 7—HOURLY REVENUES AND OVERBILLING POTENTIAL FACTORS

	Flagged physicians	Unflagged physicians
Reported hourly revenue (\$)	116.325 (1.439)	157.434 (0.172)
Predicted hourly revenue (\$)	138.255 (0.688)	159.104 (0.105)
Overbilling potential factor 1	1.907 (0.033)	0.575 (0.001)
Overbilling potential factor 2	5.978 (0.162)	1.150 (0.003)
Observations	4,412	187,654

Notes: The table compares the hourly revenues and OPFs (defined in equations (4) and (5)) 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.

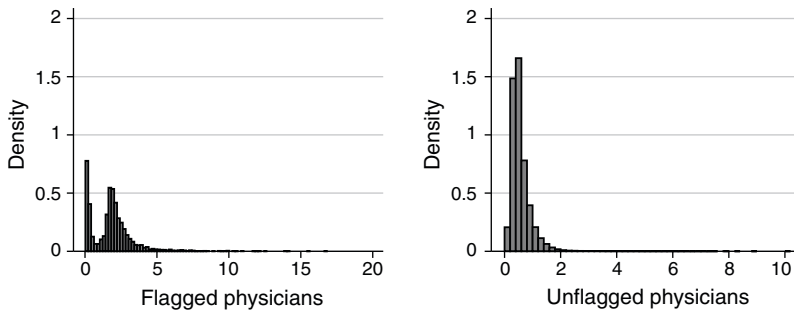
Table 7 summarizes the reported and predicted hourly revenues and the two OPFs. For flagged physicians, the reported hourly revenue is \$22, or 19 percent, less than the predicted revenue; but for unflagged physicians, reported and predicted hourly revenues are almost identical. Flagged physicians have an average OPF1 of 1.907, meaning that the total revenue from Medicare Part B FFS beneficiaries is almost twice as high as that of an unflagged peer with identical observable characteristics, *assuming their actual hours worked are identical* (8 hours per day for 365 days). Similarly, the average OPF2 for flagged physicians is 5.978, suggesting that the reported hours could be six times as much as the likely true hours worked.

Panels A and C in Figure 5 plot the distribution of OPF1 and OPF2, respectively, among the flagged and unflagged physicians. Note that, despite the heterogeneity within flagged and unflagged physicians, the distributions under both OPFs for the flagged physicians represent a substantial rightward shift of those for the unflagged physicians. In panel B we present the scatter plots of the predicted hourly revenue and the reported hourly revenue. It is clear that the majority of flagged physicians have lower reported hourly revenue than the predicted hourly revenue, whereas the opposite is true for unflagged physicians.

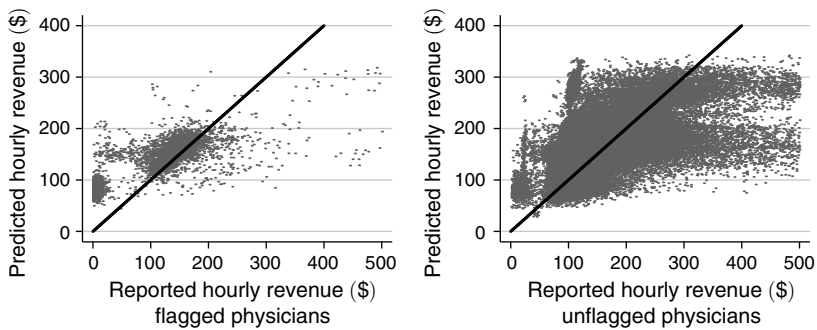
E. Coding Decisions and Fee Differentials

We now test whether coding decisions respond to financial incentives provided by different levels of service codes, and examine how they differ between flagged and unflagged physicians. In particular, we are interested in physicians' choice of code *intensities* conditional on filing a code from a given cluster. Using the code cluster presented in Table 1 as an example, we would like to know why a physician bills code 99203 more often than 99202 when a service in this code cluster—office or other outpatient visits for new patients—is furnished.

Panel A. Distribution of OPF1



Panel B. Reported and predicted hourly revenues



Panel C. Distribution of OPF2

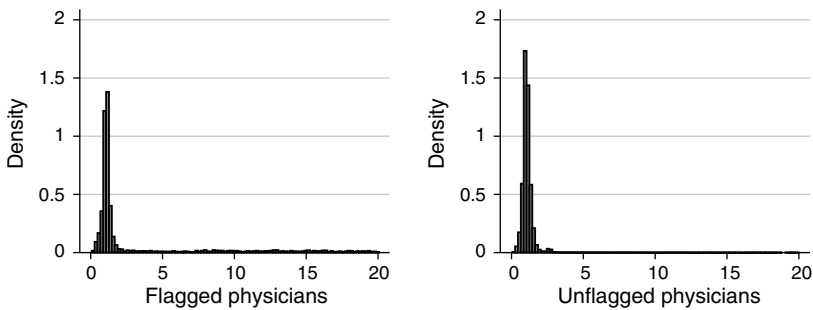


FIGURE 5. OVERBILLING POTENTIAL FACTORS

Notes: Panel A shows the distributions of OPF1 for flagged (left) and unflagged (right) physicians. Panel B shows the 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 effects, against reported hourly revenues (on the horizontal axis) for flagged and unflagged physicians. The thick solid line is the 45-degree line. Panel C shows the distributions of OPF2 for flagged (left) and unflagged (right) physicians. The bin widths in all four histograms are 0.2.

For this purpose, we analyze the physicians' coding decisions by K , the number of intensity levels in a code cluster, for $K = 3, 4, 5$. For each $K \in \{3, 4, 5\}$, we use the following baseline regression specification:

$$(6) \quad Y_{ijt} = \beta_0 + \beta_1 \text{Flagged}_{it} + \beta_{22} 1\{\text{Intensity}_j = 2\} + \cdots + \beta_{2K} 1\{\text{Intensity}_j = K\} \\ + \beta_{32} \text{Flagged}_{it} \times 1\{\text{Intensity}_j = 2\} + \cdots + \beta_{3K} \text{Flagged}_{it} \\ \times 1\{\text{Intensity}_j = K\} + \alpha_{\text{HRR}} + \eta_J + \phi_t + \varepsilon_{ijt},$$

where Y_{ijt} is the number of times physician i filed code j in year t ; Flagged_{it} is an indicator for whether physician i is flagged in year t ; $1\{\text{Intensity}_j = 2\}$ is an indicator for code j having intensity 2 in its cluster; $1\{\text{Intensity}_j = K\}$ is defined likewise, with K , the highest intensity level, being 3, 4, or 5 depending on the cluster; α_{HRR} is the HRR fixed effect; η_J is the code cluster fixed effect, where J is the cluster that j belongs to; ϕ_t is the year fixed effect; and ε_{ijt} is the error which will be clustered at the physician level.

Our primary interest is in coefficients $(\beta_{32}, \dots, \beta_{3K})$, which capture the *excess* tendency of flagged physicians to file codes at varying intensity levels relative to their unflagged peers. A positive β_{3K} , for example, indicates that flagged physicians bill more highest-intensity codes than unflagged physicians. And because a higher code intensity translates into more work RVU and thus greater Medicare reimbursement, this is consistent with flagged physicians responding to financial incentives when choosing which code within a cluster to bill. Note that being flagged does not immediately imply having a different *distribution* of code intensities. One could be flagged for billing many codes and/or for higher-intensity codes. The empirical specification allows for both effects and does not assume a priori the dominance of either effect: β_1 captures the volume effect, whereas $(\beta_{32}, \dots, \beta_{3K})$ capture the flagged physicians' excess tendency to bill for codes of different intensities.

Columns 1–3 in Table 8 report the estimation results on the subsample of codes in clusters with $K = 3, 4, 5$, respectively. Taking $K = 3$, i.e., clusters of codes with three different intensity levels, as an example, a flagged physician files 369 more codes with intensity level 2, and 128 more codes with intensity level 3 than an unflagged physician with identical observables. The same is true for other code clusters, although the estimates are much noisier for those in $K = 4$ clusters due to a small sample size. These results show that codes chosen by flagged physicians strongly tilt toward higher intensities to an extent that is hard to explain by service specialization or exposure to different markets.

In column 4 of Table 8, we pool codes from all clusters and reclassify the intensities to three levels, *low*, *middle*, and *high* as specified in Table 9. The regression results show that flagged physicians tend to file more mid-intensity codes than their unflagged peers, but not so much for high-intensity codes. This seems to contradict the hypothesis that financial incentives affect the coding decisions of flagged physicians. However, we find very different patterns when we run the baseline regression separately on the two subsamples where the marginal increase in Work RVU between two adjacent levels is *below* average (column 5) and *above* average (column 6). For example, suppose that Work RVUs increase by 100 percent per intensity level on

TABLE 8—BILLING PATTERNS AND CODE INTENSITY LEVEL

	$K = 3$ (1)	$K = 4$ (2)	$K = 5$ (3)	All K (4)	All K and below average (5)	All K and above average (6)
Flagged	246.9 [55.52]	542.9 [209.9]	37.36 [18.46]	194.5 [16.63]	215.7 [26.50]	156.8 [18.84]
Intensity = 2	242.2 [2.796]	169.7 [9.993]	11.87 [3.230]			
Intensity = 3	130.0 [2.430]	148.6 [10.54]	240.4 [3.365]			
Intensity = 4		-80.39 [10.62]	235.5 [3.181]			
Intensity = 5			33.92 [3.036]			
Flagged \times (intensity = 2)	369.2 [69.39]	155.4 [274.5]	91.08 [22.25]			
Flagged \times (intensity = 3)	128.2 [63.72]	90.05 [267.4]	143.4 [25.23]			
Flagged \times (intensity = 4)		257.8 [272.4]	57.47 [23.03]			
Flagged \times (intensity = 5)			0.115 [21.46]			
Mid-intensity				240.0 [1.761]	21.48 [1.302]	342.5 [2.837]
High-intensity				154.8 [1.509]	36.74 [1.277]	186.3 [2.499]
Flagged \times mid-intensity				91.72 [23.25]	-114.4 [26.77]	231.0 [32.79]
Flagged \times high-intensity				-21.58 [20.17]	-112.4 [26.33]	76.75 [28.22]
HRR	Yes	Yes	Yes	Yes	Yes	Yes
Code cluster	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.194	0.057	0.172	0.158	0.164	0.081
Observations	399,907	53,521	561,657	1,015,085	508,478	506,607

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 reclassify intensities to low, middle, and high as specified in Table 9. Columns 5 and 6 use the subsample of codes with below- and above-average marginal increase 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.

average, then codes 99201 through 99205 enter the regression in column 5 instead of 6 because the Work RVU increase between any two adjacent intensities in that code cluster is less than 100 percent. We find that flagged physicians do not tend to file more higher-intensity codes with below-average Work RVU increments; but they do so for codes with above-average Work RVU increments. This shows that flagged physicians do not simply overfile all codes with higher intensities. Instead, data suggest that the coding patterns are consistent with a hypothesis that flagged physicians respond to financial incentives: recall that work RVUs are closely related to Medicare reimbursements. In particular, for codes where the marginal revenue

TABLE 9—RECLASSIFICATION OF CODE INTENSITIES

	Original intensity			Reclassified intensity
	$K = 3$	$K = 4$	$K = 5$	
1	1	1	1 or 2	Low
2		2 or 3	3	Middle
3		4	4 or 5	High

Note: This table shows how the original code intensities (shown in columns $K = 3$, $K = 4$, and $K = 5$) are reclassified into three levels, low, middle, and high.

from upcoding is relatively low, flagged physicians actually file fewer mid- and high-intensity codes than their peers; but they do file more mid- and high-intensity codes when the marginal gain in revenue from upcoding is relatively high.²⁵

The regression analysis above focuses on upcoding within service clusters. There could potentially be other ways of overbilling that are not captured in these results. Physicians could bill for more of a *given* service code than it was actually provided, regardless of its intensity; or could upcode *across* code clusters by misreporting the *type of service* provided (e.g., office visits of new patients, which are paid more versus office visits of established patients, which are paid less). There is no shortage of such overbilling practices according to Department of Health and Human Services (2015). For this reason, the regression results above are likely to be lower-bound estimates of the extent to which flagged physician are potentially overbilling.

IV. Supplemental Results from External Data

Before concluding this paper, we corroborate our findings using two external datasets. Doing so both serves as a sanity check for our approach toward detecting potential Medicare overbilling, and at the same time points to possible directions in which our approach may improve existing ones.

A. Physician Working Patterns in the National Ambulatory Medical Care Survey (NAMCS) Data

The NAMCS, by the National Center for Health Statistics, Centers for Disease Control and Prevention, provides a nationally representative sample of office-based physicians. Each sampled physician is randomly assigned a week for which detailed visit-level data are collected. Given that the CMS Medicare Part B FFS sample covers the vast majority of physicians in the US, the two should be fairly comparable.²⁶

²⁵Note how conditioning on clusters helps mitigate the impact of confounding factors. For example, one might be concerned that the results in columns 5 and 6 are mechanically driven by the correlation between service fee and time, i.e., flagged physician with long hours naturally billed more high-price codes with high intensities and relatively fewer low-price ones, which is observationally similar to deliberately selecting codes with larger-than-average fee differentials that we hypothesized. But because the regressions already include cluster fixed effects, all results are conditional on the code cluster, which represents a given service. Therefore, the long hours per se do not imply billing higher-intensity codes *within a given cluster*, and would not mechanically lead to flagged physicians' inclination to bill certain codes depending on the fee differentials.

²⁶We discuss the comparability of the two samples in more detail in the online Appendix.

TABLE 10—SHARE OF MEDICARE SERVICES

	NAMCS
Share of Medicare patients	0.259 (0.006)
Share of Medicare services	0.260 (0.006)
Share of time spent with Medicare patients	0.261 (0.006)
Share of revenue from Medicare	0.297 (0.005)
Number of unique physicians	3,583

Note: Standard errors of the mean estimator are reported in parentheses.

We first examine the self-reported fraction of Medicare services by physicians sampled in NAMCS. In Table 10, Medicare patients are those whose primary payer is Medicare; fraction of Medicare patients is the weighted average of the fractions of such patients among all the sampled patients of the reporting physician; fraction of Medicare services and fraction of time spent with Medicare patients are defined similarly using the number of services and the recorded time physicians spent with Medicare patients; and fraction of revenue from Medicare is imputed from a categorical variable describing the fraction of the reporting physician's revenue from Medicare payments, with the categories being 0–25, 26–50, 51–75, and 76–100 percent. The figure reported in this table is calculated under the assumption that the actual fractions are uniformly distributed within each bin.

Next, we look into physician hours worked. Ideally, we would like to use NAMCS to calculate the self-reported total hours worked in the sampled week for each surveyed physician, who are asked to document the time they spent on patients during each visit. However, NAMCS does not sample all the visits within the chosen week, and the sampling rate varies from 100 percent for very small practices and 10 percent for very large practices. Because we don't have information on practice sizes (except for whether the physicians work in a solo or group practice), we cannot infer the total number of visits from the NAMCS sample. That said, the *maximum sampled number of hours* spent on Medicare patients per week is 15.17 for physicians working in a solo practice (all of whose visits are sampled), and 49.82 for those working in group practices (the size of which is unknown).

The comparison above shows that Medicare services typically account for about one-third of a physician's entire workload, and take far fewer hours than the 100-hour threshold we used to flag potential overbilling (at least for physicians working in solo practices). This again supports our view that the approach we develop to flag physicians for potential overbilling is likely to be conservative.

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

Finally, we relate the HCPCS code CFIs that we constructed in Section IIIC with the findings from the CMS CERT program. CERT draws a "statistically valid random sample of claims" (about 50,000) every year, requests documentation from the

filing providers, and hires medical review professionals to determine whether payments to these claims are proper or not based on their documentation. The reviewers can disapprove improper payments to claims that have insufficient documentation, questionable medical necessity, incorrect coding, or for other reasons.²⁷ The disapproval rate for Medicare Part B claims, calculated as the percentage disapproved in all sampled Part B claims, is 18.93 percent in 2012, and CMS reports Part B improper payment rate—the percentage of Medicare *dollars* paid incorrectly—for 2012 to be 12.1 percent (see Centers for Medicare and Medicaid Services 2015a).²⁸

Applying the same idea as the code flag index described in Section IIIC, we can use the prevalence of a code j among the disapproved claims and its prevalence among the approved claims to construct a code disapproval index (CDI) for the 1,621 HCPCS codes reviewed by CERT:

$$(7) \quad \text{CDI}_j = \frac{100 \times \Pr(j | \text{disapproved claims})}{\Pr(j | \text{disapproved claims}) + \Pr(j | \text{approved claims})},$$

where $\Pr(j | \text{disapproved claims})$ is the fraction of claims for HCPCS code j among all disapproved claims; and $\Pr(j | \text{approved claims})$ is the fraction of claims for HCPCS code j among all approved claims. The CDI also ranges from 0 to 100. If HCPCS code j appears only among disapproved claims, then its CDI takes on a value of 100. If a code only appears among approved claims, then its CDI is equal to 0.

We should note that the CDI as calculated from the pools of approved and disapproved claims in CERT data is *not directly comparable* to the CFI we calculated in Section IIIC. CDI is based on the prevalence of a code j among the disapproved claims relative to its prevalence among the approved claims, while CFI is based on the prevalence of a code j among the flagged physicians relative to its prevalence among the unflagged physicians. Note that we flag physicians based on whether the hours worked implied by their claims are implausibly long, but we do *not* take a stand on whether or not *any particular claim* is suspicious. On the other hand, CERT program is examining whether particular claims are legitimate, but does not take into account of the overall billings of the physicians. Thus, we believe that the CFI we construct and CDI calculated from the CERT data are complementary.

In Figure 6 we compare CDIs and CFIs for the 1,621 HCPCS codes that appear in the CERT data, where each HCPCS code is represented by a circle, with the radius proportional to its total Medicare reimbursement in our CMS sample.²⁹ Codes falling into the southwest quarter of the plane have both CFIs and CDIs below 50. On the other hand, codes in the northeast quarter are those with both indices above 50, and are thus more frequently associated with (potentially) inappropriate billing practice. These two quarters are cases where our flagging approach and CERT review results agree. On the other hand, codes in the southeast quarter are those with high CDIs from CERT but low CFIs in our sample. Similarly, codes in the northwest

²⁷The disapproved claims will have any overpayment recouped, but do not receive other sanctions.

²⁸Improper payment defined by CMS includes both overpayment and underpayment. The latter contributes only 0.2 percentage points to the 12.1 percent improper payment rate as reported by Centers for Medicare and Medicaid Services (2015a).

²⁹The HCPCS codes not reviewed by CERT make up 6.2 percent of total reimbursement in our CMS sample.

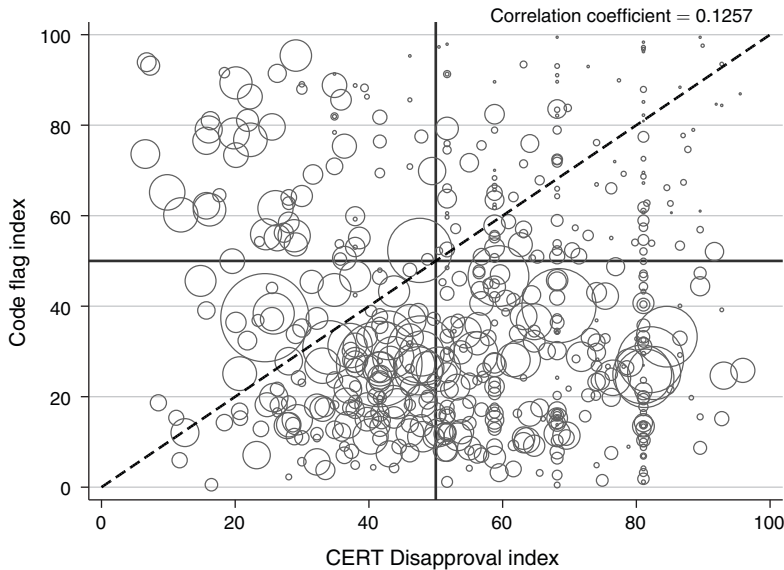


FIGURE 6. HCPCS CODE FLAG INDEX AND CERT CODE DISAPPROVAL INDEX

Notes: The horizontal axis shows the CERT code disapproval index, defined in equation (7). The vertical axis shows the CFI, defined in equation (3). 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 1,621 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.

quarter are those with high CFIs but low CERT CDIs. The overall unweighted correlation between the two indices is 0.1257.

The comparison between CDI and CFI suggests that our approach to constructing a code flag index based on flagged and unflagged physicians could potentially contribute to existing auditing methods. Reformulating sampling strategies to focus more on HCPCS codes with high CFIs, especially those where CFI and CERT CDI differ substantially, may help better detect and deter inappropriate billing with limited regulatory resources. The CFI we construct can help screen codes that are more likely to be associated with potential overbilling.

V. Conclusion

In this paper, we propose and implement a novel approach to detect potential overbilling in Medicare reimbursement based on the simple idea that all physicians have a fixed time budget in a given period (a calendar year, for example) and the services claimed for reimbursement require time to complete. We construct the implied hours worked at the individual physician level based on service codes submitted to Medicare. We flag physicians as potentially overbilling based on whether the implied hours worked are implausibly long. One caveat, however, is that the estimated *level* of potential overbilling might be biased downward due to data limitations, especially the exclusion of services submitted to payers other than Medicare Part B FFS. Nonetheless, our method for detecting potential overbilling is still useful with at least three advantages relative to the existing methods. First, it imposes

minimal data requirements, and is easy to implement, automate, and update over time. Second, it mitigates the impact of confounding factors in the detection of overbilling such as selection and physician heterogeneity, because all physicians face the same time constraint regardless of their patient pool or practice patterns. Finally, it allows users to freely choose the level of stringency when flagging potential overbilling. For example, one could use a different threshold of weekly hours worked or, if higher-frequency data are available, flag physicians based on claims filed in a quarter, a month, or even a week, in which case there is less intertemporal smoothing than is permitted in our sample. Interestingly, we also find suggestive evidence that the coding patterns of the flagged physicians seem to be responsive to financial incentives: within code clusters with different levels of service intensity, they tend to submit more higher-intensity service codes than unflagged physicians; moreover, they are more likely to do so if the marginal revenue from submitting mid- or high-intensity codes is relatively high.

Overbilling arises from the strong information asymmetry between providers, patients, payers, and regulators, exacerbated by the financial incentive of doing so under an FFS payment regime. We believe that efficient and effective auditing, among other approaches, is crucial to curbing, if not curing, the overbilling problem. Our approach provides a quick and easy tool for detecting potential overbilling, but we would like to emphasize that it does not provide definite evidence for fraudulent coding, nor does it substitute existing methods based on auditing. It is also silent on whether the current Physician Fee Schedule is compensating the physicians fairly. Rather, we take the existing regulations as given, and view our approach as a useful screening tool to identify individual physicians, specialties, or HCPCS codes whose billing patterns are highly consistent with overbilling and are hard to reconcile using observables. For example, the HCPCS code CFIs suggest that certain codes are disproportionately associated with flagged physicians, and thus may need more auditing attention. This can help improve the efficiency in the allocation of limited regulatory resources.

Another possibility is to focus on correcting the distorted incentives of FFS, which would require more changes to the existing system. Replacing FFS with capitation seems appealing, but could also introduce new distortions. For example, Geruso and Layton (2015) find that providers tend to upcode the *diagnoses* of patients with Medicare Advantage plans to game the risk-adjustment process. Effective auditing within the FFS can not only detect error and fraud that already occurred, but also have a long-term deterrence effect.

REFERENCES

- Adams, Diane L., Helen Norman, and Valentine J. Burroughs.** 2002. "Addressing Medical Coding and Billing Part II: A Strategy for Achieving Compliance. A Risk Management Approach for Reducing Coding and Billing Errors." *Journal of the National Medical Association* 94 (6): 430–47.
- Adler-Milstein, Julia, and Ashish K. Jha.** 2014. "No Evidence Found That Hospitals Are Using New Electronic Health Records to Increase Medicare Reimbursements." *Health Affairs* 33 (7): 1271–77.
- Angeles, January, and Edwin Park.** 2009. "Upcoding" *Problem Exacerbates Overpayments to Medicare Advantage Plans*. Washington, DC: Center on Budget and Policy Priorities.
- Bastani, Hamsa, Joel Goh, and Mohsen Bayati.** 2016. "Evidence of Strategic Behavior in Medicare Claims Reporting." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2630454 (accessed February 13, 2015).

- Becker, David, Daniel Kessler, and Mark McClellan. 2005. "Detecting Medicare Abuse." *Journal of Health Economics* 24 (1): 189–210.
- Bergman, Jonathan, Christopher S. Saigal, and Mark S. Litwin. 2015. "Service Intensity and Physician Income: Conclusions from Medicare's Physician Data Release." *JAMA Internal Medicine* 175 (2): 297–99.
- Bowblis, John R., and Christopher S. Brunt. 2014. "Medicare Skilled Nursing Facility Reimbursement and Upcoding." *Health Economics* 23 (7): 821–40.
- Brunt, Christopher S. 2011. "CPT Fee Differentials and Visit Upcoding under Medicare Part B." *Health Economics* 20 (7): 831–41.
- Centers for Medicare and Medicaid Services. 2013. "Medicare Physician Fee Schedule: Payment System Fact Sheet Series." Baltimore: US Department of Health and Human Services.
- Centers for Medicare and Medicaid Services. 2014. "How to Use the Searchable Medicare Physician Fee Schedule (MPFS)." Baltimore: US Department of Health and Human Services.
- Centers for Medicare and Medicaid Services. 2015a. "Medicare Fee-For-Service 2014 Improper Payments Report." Baltimore: US Department of Health and Human Services.
- Centers for Medicare and Medicaid Services. 2015b. "Medicare Learning Network: Evaluation and Management Services." Baltimore: US Department of Health and Human Services.
- Chan, David, and Michael Dickstein. 2016. "Affiliation and Price-Setting by Committee in Medicare." <https://ashecon.confex.com/ashecon/2016/webprogram/Paper4984.html> (accessed June 17, 2016).
- Clair, Brandon, and Parul Goyal. 2015. "What Does Medicare Pay Rhinologists? An Analysis of Medicare Payment Data." *International Forum of Allergy and Rhinology* 5 (6): 481–86.
- Congressional Budget Office. 2014. *The 2014 Long-Term Budget Outlook*. Washington, DC: US Congress.
- Cutler, David, Jonathan Skinner, Ariel Dora Stern, and David Wennberg. 2015. "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending." Harvard Business School Working Paper 15-090.
- Dafny, Leemore, and David Dranove. 2009. "Regulatory Exploitation and Management Changes: Upcoding in the Hospital Industry." *Journal of Law and Economics* 52 (2): 223–50.
- Dusetzina, Stacie B., Ethan Basch, and Nancy L. Keating. 2015. "For Uninsured Cancer Patients, Outpatient Charges Can Be Costly, Putting Treatments Out of Reach." *Health Affairs* 34 (4): 584–91.
- Fang, Hanming, and Qing Gong. 2017. "Detecting Potential Overbilling in Medicare Reimbursement via Hours Worked: Dataset." *American Economic Review*. <https://doi.org/10.1257/aer.20160349>.
- Fryer, Roland G., Jr., and Steven D. Levitt. 2004. "The Causes and Consequences of Distinctively Black Names." *Quarterly Journal of Economics* 119 (3): 767–805.
- Gabbert, W., K. H. Kachur, and T. Whitehead. 2012. *Current Procedural Coding Expert*. Salt Lake City: Ingenix.
- Geruso, Michael, and Timothy Layton. 2015. "Upcoding: Evidence from Medicare on Squishy Risk Adjustment." National Bureau of Economic Research Working Paper 21222.
- Gino, Francesca, Shahar Ayal, and Dan Ariely. 2009. "Contagion and Differentiation in Unethical Behavior: The Effect of One Bad Apple on the Barrel." *Psychological Science* 20 (3): 393–98.
- Harewood, Gavin C., Gary Foley, and Zarah Farnes. 2014. "Pricing Practices of Gastroenterologists in New York." *Clinical Gastroenterology and Hepatology* 12 (11): 1953–55.
- Institute of Medicine of the National Academies. 2009. *Resident Duty Hours: Enhancing Sleep, Supervision, and Safety*. Washington, DC: National Academies Press.
- Ip, Ivan K., Ali S. Raja, Steven E. Seltzer, Atul A. Gawande, Karen E. Joynt, and Ramin Khorasani. 2015. "Use of Public Data to Target Variation in Providers' Use of CT and MR Imaging among Medicare Beneficiaries." *Radiology* 275 (3): 718–24.
- Jones, Lyell K., Karolina Craft, and Joseph V. Fritz. 2015. "Medicare Payment Transparency: Implications for Neurologists." *Neurology: Clinical Practice* 6 (1): 73–80.
- Kaiser Family Foundation. 2015. *The Facts on Medicare Spending and Financing*. Menlo Park: Kaiser Family Foundation.
- Ko, Joan S., Heather Chalfin, Bruce J. Trock, Zhaoyong Feng, Elizabeth Humphreys, Sung-Woo Park, H. Ballentine Carter, Kevin D. Frick, and Misop Han. 2015. "Variability in Medicare Utilization and Payment among Urologists." *Journal of Urology* 85 (5): 1045–51.
- Lapps, Joshua, Bradley Flansbaum, Luci Leykum, Josh Boswell, and Luigi Haines. 2016. "Updating Threshold-Based Identification of Hospitalists in 2012 Medicare Pay Data." *Journal of Hospital Medicine* 11 (1): 45–47.
- Lorence, Daniel P., and Amanda Spink. 2002. "Regional Variation in Medical Systems Data: Influences on Upcoding." *Journal of Medical Systems* 26 (5): 369–81.

- Menger, Richard P., Michael E. Wolf, Sunil Kukreja, Anthony Sin, and Anil Nanda.** 2015. "Medicare Payment Data for Spine Reimbursement: Important but Flawed Data for Evaluating Utilization of Resources." *Surgical Neurology International* 6 (S14): S391–97.
- Merrell, Katie, Claudia Schur, Stephen Zuckerman, and Robert Berenson.** 2016. "Physician Service Time in the Medicare Physician Fee Schedule." <https://ashecon.confex.com/ashecon/2016/webprogram/Paper4112.html> (accessed June 15, 2016).
- O’Gara, Patrick T.** 2014. "Caution Advised: Medicare’s Physician-Payment Data Release." *New England Journal of Medicine* 371: 101–03.
- Physicians Foundation.** 2012. *A Survey of America’s Physicians: Practice Patterns and Perspectives*. The Physicians Foundation.
- Rosenberg, Marjorie A., Dennis G. Fryback, and David A. Katz.** 2000. "A Statistical Model to Detect DRG Upcoding." *Health Services & Outcomes Research Methodology* 1 (3): 233–52.
- Schmajuk, G., K. J. Bozic, and J. Yazdany.** 2014. "Using Medicare Data to Understand Low-Value Health Care: The Case of Intra-Articular Hyaluronic Acid Injections." *JAMA Internal Medicine* 174 (10): 1702–04.
- Silverman, Elaine, and Jonathan Skinner.** 2004. "Medicare Upcoding and Hospital Ownership." *Journal of Health Economics* 23 (2): 369–89.
- Skolarus, Lesli E., James F. Burke, Brian C. Callaghan, Amanda Becker, and Kevin A. Kerber.** 2015. "Medicare Payments to the Neurology Workforce in 2012." *Neurology* 84 (17): 1796–1802.
- Skolasky, Richard L., and Lee H. Riley III.** 2015. "Medicare Charges and Payments for Cervical Spine Surgery: Association with Hospital Characteristics." *Spine* 40 (16): E936–42.
- Sutphin, P. D., A. Ding, S. Toomay, S. P. Reis, A. K. Pillai, S. P. Kalva, and S. L. Hsu.** 2014. "Interventional Radiologist Billing Practices to Medicare: Procedures Performed and Geographic Variation." *Journal of Vascular and Interventional Radiology* 26 (2): S173–74.
- US Department of Health and Human Services.** 2004. *HIPAA Administrative Simplification: Standard Unique Health Identifier for Health Care Providers; Final Rule*. Washington, DC: Office of Inspector General.
- US Department of Health and Human Services.** 2015. *A Roadmap for New Physicians: Avoiding Medicare and Medicaid Fraud and Abuse*. Washington, DC: Office of Inspector General.
- Wennberg, John E., and Megan M. Cooper.** 1996. *The Dartmouth Atlas of Health Care*. Chicago: American Hospital Publishing.
- Zeltzer, Dan.** 2016. "Gender Homophily in Referral Networks: Consequences for the Medicare Physician Pay Gap." https://dl.dropboxusercontent.com/content_link/ynHHZx2NESNRpUjlrrogWiBdah23OoBX50DiBnlaevHKYAfe4KjZCJ5JYD1hWU6XS/file (accessed July 26, 2016).
- Zuckerman, Stephen, Robert Berenson, Katie Merrell, Tyler Oberlander, Nancy McCall, Rebecca Lewis, Sue Mitchell, and Madhu Shrestha.** 2014. *Development of a Model for the Valuation of Work Relative Value Units: Objective Service Time Task Status Report*. Baltimore: Centers for Medicare and Medicaid Services.