Medicare Home Health Fraud: How Much, Where, and Who?

Liran Einav, Amy Finkelstein, Yunan Ji, Neale Mahoney, and Gideon Moore*

May 25, 2026

Abstract

How much fraud is there in Medicare and who commits it? We provide an answer for Medicare home health, a setting widely considered especially rife with fraud. We define a home health agency (HHA) as fraudulent if it was prosecuted by a federal strike force. Combining Medicare claims data on all HHAs with hand-collected prosecution records from the nine federal judicial districts where strike forces operated between 2009 and 2013, we train a machine learning model to predict, out of sample, the probability that each HHA in the remaining 85 districts would have been prosecuted had a strike force been present. We estimate that in 2008, 3.4% of Medicare home health spending — about $520 million — was billed by fraudulent HHAs. The strike forces were well-targeted: their nine districts contained only 40% of home health spending but 65% of fraudulent spending. Fraudulent HHAs display intuitive characteristics: they are more likely to rely on extremely high-volume referring physicians, to exhibit unusually uniform patterns of care, and to serve healthier-than-average patients.

*Einav: Stanford and NBER, leinav@stanford.edu; Finkelstein: MIT and NBER, afink@mit.edu; Ji: Georgetown and NBER, yunan.ji@georgetown.edu; Mahoney: Stanford and NBER, nmahoney@stanford.edu; Moore: Stanford, gsmoore@stanford.edu. We are grateful to Abby Joseph, Ken Lin, and Steven Shi for excellent research assistance. We thank Stephen Lee, Jetson Leder-Luis, Han Hong, Jon Skinner, and Jim Poterba for useful conversations. We thank the Laura and John Arnold Foundation for financial support.

1 Introduction

Health care accounts for almost one-fifth of the US economy, with about half of that spending financed by taxpayers. There is broad consensus that much of this spending is unnecessary, with some estimates suggesting that as much as a quarter of it could be eliminated without harm to patients (Orszag 2009; Shrank et al. 2019). Identifying how to do so, however, has proved challenging (e.g., Cutler 2010; Doyle et al. 2015). Reducing health care fraud is therefore a particularly appealing target, since it suggests a clear source of health care spending that can be reduced without harm to patients. Not surprisingly, therefore, policymakers and pundits often invoke the potential savings from combating health care fraud in their plans to reform the sector.

How much health care fraud is out there? The Government Accountability Office (GAO) defines health care fraud as intentionally deceptive delivery or billing of medical care, such as billing for services or supplies that are not actually provided. It has estimated that in 2016, about 8% of total health-care spending in CMS-managed health care programs, or roughly $95 billion, fell into the category of "waste, fraud, and abuse." But this category runs the gamut from waste (such as over-use of low-value care), to abuse (such as up-coding and other aggressive billing practices), to outright fraud (General Accounting Office 2017; Leder-Luis and Malani 2025). How much of this 8% is actually fraud? Surprisingly little academic work has tried to estimate the amount or nature of health-care fraud. This paper provides an estimate, for one specific setting.

We focus on Medicare-financed home health. This service is large and growing; in 2019, one-in-five individuals aged 85 and over received Medicare-financed home health care.[1] It is also considered particularly susceptible to fraud (MedPAC 2017, Department of Health and Human Services (2018)). Eligibility is difficult to verify, care is delivered at home where there is less monitoring, patients have no cost sharing, and compared to other types of medical care, the service has limited medical risks and may provide direct utility to healthy individuals. We therefore suspect that the fraudulent share of Medicare home health spending is higher than for other Medicare-covered services. To estimate the share of Medicare home health spending that is fraudulent, we use the geographically targeted strike forces deployed by the federal government in response to a 115% increase in Medicare home health spending between 2000 and 2008.[2] Between 2009 and 2013, these strike forces were launched in nine of the 94 federal judicial districts, starting with Miami-Dade County in Southern Florida.[3]

The strike forces' prosecutions provide our operational measure of home health fraud. There is a large gray area of activities that could plausibly be classified as fraud or as "waste or abuse," and the distinction is ultimately legal rather than economic. We therefore define fraud based on what strike

[1] Authors' calculation from Medicare data.

[2] Authors' calculations from Medicare home health claims data.

[3] Two other home health strike forces were subsequently opened in 2018 in the District of New Jersey (which contains Newark) and the Eastern District of Pennsylvania (which contains Philadelphia). As of the end of 2025, there had been no additional home health strike force launches.

forces successfully prosecuted. Specifically, we compile a data set of all 118 prosecutions of home health agencies (HHAs) brought in the nine strike-force districts between 2009 and 2019, restricting attention to HHAs that existed in 2008 (the start of the strike forces). Because prosecutions are directed against agencies rather than specific claims, we classify all 2008 claims from an HHA that was prosecuted as fraudulent. Our approach may overstate fraud at prosecuted HHAs since some share of their activity may be legitimate. Our approach may also understate fraud since the high cost of prosecutions and high burden of proof mean that some “known” cases of fraud are not prosecuted. Our definition of fraud thus reflects what was successfully prosecuted, given the current legal standards and typical strike force resources.

The strike forces also provide our empirical approach for estimating the amount of home health fraud nationally. The observed prosecutions in the nine strike-force districts allow us to build a machine learning model that predicts prosecution. The model uses 2008 HHA characteristics along with rich 2003–2008 Medicare claims data describing the services the HHAs provide, the patients they serve, and the physicians who refer to them. We then use the model to predict, out of sample, the probability that each HHA that existed in 2008 in the remaining 85 districts would have been prosecuted had a strike force (counterfactually) entered between 2009 and 2013. These predictions yield two outputs: an estimate of the share of home health spending billed by fraudulent HHAs nationally, and a characterization of the features that distinguish fraudulent HHAs.

We have two main sets of findings. The first concerns the *amount* of fraud. We estimate that in 2008, prior to the entry of the strike forces, 3.4% of national Medicare home health spending – about $520 million – was billed by fraudulent HHAs. The fraudulent share for other Medicare-covered services is likely lower than our estimates for home health, since Medicare home health is considered particularly fraud-prone. Even still, applying our 3.4% estimate to all Medicare spending would imply roughly $34 billion in Medicare fraud in 2023 (Commission et al. 2025), which would be at the low end of the 3-10% range commonly cited in prior policy estimates (National Health Care Anti-Fraud Association; Senior Medicare Patrol). Our estimate is also substantially below the only prior academic estimate of Medicare home health fraud we are aware of, which places the share at 6% to 24% (Lee and Skinner 2021); however they consider “waste and abuse” as well, without restricting attention to fraud as we define it. Put differently, our findings suggest that prosecuting Medicare fraud could reduce at most two-fifths – and likely considerably less – of the 8% of Medicare spending that the GAO characterizes as “waste, fraud, and abuse” (General Accounting Office 2017).

Our second set of findings concerns the geography and characteristics of fraudulent HHAs. Fraud is highly geographically concentrated and the strike forces were well-targeted. Although the nine strike-force districts accounted for only 39% of home health spending in 2008, they accounted for 65% of the home health spending by fraudulent HHAs. On average, fraudulent HHAs accounted for 4.5% of home health spending in a strike-force district, and 2.0% in the non-strike-force districts.

These fraudulent HHAs display intuitive characteristics. For example, they are more likely to rely on extremely high-volume referring physicians, to exhibit unusually uniform patterns of care, and to serve healthier-than-average patients.

Our findings relate to two adjacent literatures. First, there is a large literature on unnecessary health care spending and how to reduce it. The consensus is that the US health care sector contains enormous waste, but that there are few instruments that can reduce that waste without risking harm to patients. Researchers have focused on three main sources of unnecessary health care spending: high prices, high administrative costs, and over-use of low-value care. Each is hard to address without unintended consequences. Lowering prices may reduce the supply of or innovation in medical care (e.g., Clemens and Gottlieb 2014; Lakdawalla 2018; Alexander and Schnell 2024). Reducing administrative complexity may save direct administrative costs but allow for more over-use of low-value care (e.g., Brot-Goldberg et al. 2023; Shi 2024). And most demand-side or supply-side instruments aimed at reducing over-use of low-value care have proven to be blunt, cutting high-value and low-value care in equal measure (e.g., Brot-Goldberg et al. 2017; Cabral and Cullen 2017; Curto et al. 2019; Geruso et al. 2023). Our paper fits into a smaller literature that has identified specific policies for cutting relatively small amounts of health-care spending without harm to patients – for example, reform to long-term care hospital reimbursement (Einav et al. 2023) or the elimination of surprise billing for out-of-network emergency care (Cooper and Morton 2016; Cooper et al. 2020).[4]

Second, there is a literature on the impact of policies aimed at reducing Medicare fraud specifically. These policies include whistle-blower filings under the False Claims Act (Leder-Luis forthcoming; Howard and McCarthy 2021), restrictions on Medicare-financed ambulance use for routine dialysis transport (Eliason et al. 2021), and revenue caps and anti-fraud litigation in Medicare hospice (Gruber et al. 2023). Within home health in particular, researchers have studied the impact of a variety of anti-fraud policies – including strike forces, moratoria on the entry of new HHAs, and a cap on the share of an HHA's annual payments that could come from outlier payments (Kim and Norton 2015; Einav et al. 2025). Our paper differs in focus: rather than estimating the impact of anti-fraud policies on total health-care spending (fraudulent or not), we estimate the amount of fraud and the characteristics of providers who engage in it.[5] Our estimates can be interpreted as the amount of fraudulent Medicare home health spending that could be reduced if strike forces were implemented across the country. Our results suggest that the ability to reduce Medicare spending by prosecuting fraud may be considerably less than prior work has hypothesized.

The rest of the paper proceeds as follows. Section 2 describes our setting and Section 3 describes

[4]A list of specific policy proposals that academic researchers view as promising ways to surgically excise small amounts of unnecessary health care spending without harming patients is available here: https://onepercentsteps.com/

[5]Related to our work, O'Malley et al. (2023) study the diffusion of home health fraud across home health agencies by documenting that areas of the country that had more sharing of patients across HHAs in 2002 were associated with higher rates of subsequent home health spending growth through 2009.

our data. Section 4 lays out our empirical approach and Section 5 reports the results. The final section concludes.

2 Setting

2.1 Medicare home health

Medicare-covered home health looms large in the lives of the elderly.[6] In 2019, one-in-twelve Medicare enrollees aged 65 and over – and one-in-five of those aged 85 and over – used Medicare-financed home health.[7] That year, Medicare spent \$20 billion on home health, accounting for almost one-third of its spending on all post-acute care.[8]

Home health involves licensed medical providers providing skilled medical and rehabilitative services to patients in their homes. Eligibility for Medicare home health requires that a patient be "homebound" (i.e., leaving home requires considerable effort) and need part-time or intermittent skilled nursing or therapy. Medicare-covered home health services must be provided as part of a treatment plan where the patient is expected to improve and no longer need these services. A physician or other qualified provider must certify this need and renew the certification at least every 60 days (the statutory length of a home health "episode" during our study period), specifying the patient's condition, service plan, and expected duration of care (42 CFR 424.22 2020). Covered services include skilled nursing, physical and occupational therapy, speech-language pathology, and limited home health aide assistance (Sengupta et al. 2022). Unlike most Medicare services, home health entails no cost sharing for beneficiaries, except for durable medical equipment provided during care episodes (CMS 2022; CMS 2023). During our 2008-2019 study period, Medicare reimbursed HHAs for home health using a prospective payment system; HHAs were paid based on 60-day "episodes of care," with payment amounts adjusted for patient needs and geographic location.[9] Some episodes may receive additional outlier payments for unusually costly cases, and some episodes may receive a reduced reimbursement, known as a low-utilization payment adjustment (LUPA), when a home health agency delivers fewer visits than the minimum threshold set for a given payment period (MedPAC 2017; Medicare Learning Network (2019), Palmetto GBA (2025)).

To serve Medicare patients, an HHA must be primarily engaged in providing skilled nursing and therapeutic services, and must employ at least one physician and one registered professional nurse to oversee its operations. HHAs are required to obtain a license from the state or locality in which they operate; as licensing regulations vary by state, agencies typically confine their activities to a single state (Famakinwa 2023). To receive Medicare certification, agencies must also submit

[6] All of our discussion of "Medicare-covered" home health – as well as prosecuted fraud – pertains to Traditional Medicare, i.e. the publicly provided insurance program for Medicare enrollees; it excludes Medicare Advantage enrollees who receive their health insurance from private insurance companies reimbursed by the government.

[7] Authors' calculation from Medicare data.

[8] Authors' calculations based on MedPAC (2022).

[9] In 2019, the national base payment was roughly \$3,000 per episode.

an enrollment application, which is reviewed and approved by Medicare (CMS 2020). In 2019, the home health industry employed approximately 1.4 million workers across more than 11,000, predominantly for-profit, HHAs (BLS 2019; CDC 2020).

2.2 Concerns about Medicare home health fraud

Medicare home health has long been considered susceptible to fraud and abuse due to the nature of the benefit: Eligibility depends on physician certification that can be hard to verify, care is delivered in the home rather than in a clinical setting, services are provided without patient cost sharing, and many of the provided services (e.g., assistance with various chores and physical therapy) would also be valuable for individuals without any medical need (Department of Health and Human Services 2018). These characteristics create scope for billing for unnecessary or unprovided services, and for enrolling ineligible patients.

Concerns about fraud intensified in the 2000s as Medicare home health spending rose rapidly following an October 2000 reform that changed how Medicare reimbursed HHAs from a fee for service to a prospective payment system. While this reform was itself a response to concerns about rapidly growing Medicare home health spending – and an interim reform from 1997 to 2000 did succeed in temporarily reversing this trend (McKnight 2006) – rapid growth resumed shortly after the 2000 reform. Growth was especially high in a few regions, most notably in Miami-Dade County in Florida and in parts of Texas, where home health expenditures per beneficiary quadrupled between 2002 and 2008 (MedPAC 2010; Lee and Skinner 2021). These patterns prompted targeted enforcement actions, starting with the strike forces that we study.

2.3 Combating home health fraud: strike forces

Our analysis focuses on the establishment of inter-agency *strike forces* – a collaboration between the Department of Justice (DOJ) and the Department of Health and Human Services' Office of Inspector General (OIG) to detect and prosecute Medicare fraud in areas with high levels of suspected fraud. The first strike force was created in 2007 in the Southern District of Florida, initially focusing on durable medical equipment; it expanded its scope in 2009 to include home health. Over the subsequent decade, eleven additional strike forces were established in other federal judicial districts, selected based on claims anomalies and other intelligence (HHS & DOJ 2010; HHS & DOJ 2011).

By 2019, strike force teams operating in these districts had filed over 2,800 criminal cases involving approximately $3.5 billion in losses (Office of the Inspector General 2020).[10] These cases

[10] Out of the 12 openings of home health Task Forces from 2009 through 2025, two (in Pennsylvania and New Jersey) opened in 2018. Given that these are much later, we exclude them from the sample. The Middle and Eastern Districts of Louisiana worked extremely closely and were often referred to jointly as "South Louisiana" in official reports (HHS & DOJ 2014); we therefore count them jointly as "one district" in the context of the analysis, so that we end up with 11 active districts, 9 of which we analyze.

did not solely target Medicare home health fraud. Other areas of targeted Medicare fraud included durable medical equipment, IV infusion therapy, and opioid prescribing. Estimates based on strike force press releases suggest that slightly over one-third of strike force prosecutions over the 2002-2019 period were for home health (Einav et al. 2025).

Each strike force team combined the investigative and analytical resources of multiple agencies, and brought forward successful cases to federal district courts (HHS & DOJ 2010; HHS & DOJ 2011). Strike force efforts involved teams of criminal prosecutors, federal and state agents, licensed nurses, and local policy investigators, all working together to investigate and bring criminal or civil charges against those suspected of committing home health fraud. A rough, ballpark estimate is that each strike force spent on average $10 million per year targeting suspected home health fraud (Einav et al. 2025, Appendix A). Suspected targets were typically investigated using both traditional detective work as well as real-time analysis of Medicare billing data to identify suspect billing patterns; cases were typically handled via the US Attorney's Office although sometimes they were handled jointly with the Civil Division of the Department of Justice (HHS & DOJ 2010, HHS & DOJ 2022, OIG 2022).[11]

Because the burden of proof is high and the resource cost of prosecutions is large, prosecutors pursue only the strongest, most egregious fraud cases (CMS 2014). Consequently, virtually all cases against home health agencies pursued by the strike forces have resulted in either conviction or settlement.[12] The most commonly prosecuted forms of home-health fraud include billing for excessive or medically unnecessary care, providing services to ineligible patients, or billing for visits that were never delivered (HHS & DOJ 2010; HHS & DOJ 2012). For example, in one case, a home health patient recruiter at a Detroit-area HHA admitted that he recruited "hundreds of patients, who were neither home bound nor in need of therapy services, through the payment of cash kickbacks in exchange for their Medicare information and signature on medical documents" and that this information was then used by the HHA to bill Medicare for home health services that were never performed. However, in another successful prosecution, the charge was billing for home health services that were "not medically necessary and for services rendered to patients who were not home-bound" (HHS & DOJ 2012). In other words, prosecuted cases include both clear examples of fraud but also examples that – from an external perspective without the additional information gathered by the prosecution – might appear indistinguishable from the GAO's definition of "abuse." This underscores the inherent difficulty for a researcher in identifying what is and is not fraud, which in turn motivates our use of strike force prosecutions as an operational definition of fraud.

[11] The experience from strike force prosecutions indicated that home health fraud tended to replicate and migrate in response to enforcement pressure, a pattern that heightened regulatory suspicion of new HHAs and ultimately led to the introduction of moratoria on the entry of new HHAs in several states (CMS 2013).

[12] Authors' analysis of data on the strike force prosecutions of home health agencies described in next section.

3 Data sources

We use two primary sources of data: hand-collected data on strike force prosecutions of individual HHAs and standard Medicare administrative data.

Strike force prosecutions. We obtained information on the strike forces, including which districts they operated in and when those operations began, from the Health Care Fraud and Abuse Control Program Annual Reports for each year during our study period (e.g., HHS & DOJ 2010). Within each of the nine districts in which a strike force operated between 2009 and 2013, we identified HHAs that were prosecuted for fraud under the strike force by compiling information from three primary public enforcement data sources: the archived press releases of the Health Care Fraud Prevention and Enforcement Action Teams (HEAT), the US Department of Justice (DOJ) website, and the individual websites of each United States Attorney's Office (USAO). These press releases contain detailed records of all indictments, convictions, and sentencing announcements arising from federal health-care fraud enforcement activities (both civil and crimminal) within each strike force district. We manually checked each press release to identify those that were related to home health. For those that were, we extracted the HHA's name (if available) and matched the prosecuted HHAs to their corresponding provider identifiers in the Medicare administrative data using multiple reference sources.[13] The Appendix contains more details on how we collected and matched these data.

Through this process, we identified 217 HHAs that were prosecuted in the nine strike force districts between 2009 and 2019; we end our analysis in 2019 due to the onset of the Covid-19 pandemic. We limit our analysis to the 54.4% of these HHAs that existed in 2008, prior to the entry of the first home health strike force in 2009. This allows us to focus on the risk set of HHAs in existence in 2008 and study their behavior prior to any response to strike force entry. We then identify those that were prosecuted for fraud over the next 10 years (from 2009 to 2019). Appendix Table OA.1 contains the list of 118 prosecuted HHAs in our final data set.

Medicare administrative data. We use three administrative Medicare datasets. The first is the annual Master Beneficiary Summary File annually for 2000-2019. This file provides data on the universe of Traditional Medicare enrollees including zip code of residence, race, and gender, coverage by other insurance (i.e. months dually covered by Medicaid as well as months covered by Medicare Advantage instead of Traditional Medicare) – as well as whether she has each of 22 different comorbidities (such as asthma, depression, diabetes, or Alzheimer's). The second dataset is the annual Medicare home health claims files from a 100% sample of Traditional Medicare enrollees, also available 2000-2019. These claims contain detailed episode-level information on home health

[13] 27% of home health fraud cases we reviewed did not name a HHA; these could be cases against physicians, staffing companies, or patient recruiters who were not part of an agency. We exclude them from our analysis.

services, including the start and end dates of each 60-day home health episode, the number and type of home health visits and services provided, the physician certifying the care, and the HHA delivering the care. For our prediction exercise, we focus on data from 2003 through 2008 to characterize home health utilization patterns prior to the launch of the first home health strike force in 2009. Finally, we use the Medicare Provider of Service file to obtain characteristics of the HHAs operating in 2008, such as their year of entry and their ownership status (e.g. for-profit, non-profit, or government owned).

Descriptive statistics on strike forces. Panel (a) of Figure 1 shows the districts and timing of the nine strike force openings during our sample period (2009-2013), as well as the two strike forces targeting home health that opened in 2018 that are not part of the analysis. Strike forces were targeted to areas suspected of having high rates of fraud. These suspicions were based on intelligence data on where fraud might be occurring as well as where analysis of Medicare claims revealed aberrant billing practices (HHS & DOJ 2010, HHS & DOJ 2011). Panel (b) of Figure 1 shows abnormally high level and rate of growth in home health spending between 2000 and 2008 in the districts targeted by the strike force. As a result, in 2008 the nine strike force districts represented 22.6% of full-year equivalent Traditional Medicare enrollees in our sample but 39% of Medicare home health spending.[14] Appendix Table OA.2 shows, for each strike force, the date it was created, and the share of claims and payments at prosecuted HHAs in 2008.

4 Predicting Fraud

Empirical approach. We use data from the nine districts in which a home health strike force entered between 2009 and 2013, and within these districts we classify HHAs as fraudulent or not based on whether they were prosecuted by the strike force. Given that prosecutions occur at the HHA level,[15] our strategy is to classify fraud at the HHA level despite the fact that, naturally, even fraudulent HHAs may have some non-fraudulent component of their activities. Our data contain claims from 3,603 HHAs that in 2008 operated in districts with strike forces, of which 118 (3.3%) were prosecuted during our analysis period (2009-2019).[16]

For our analysis, we focus on data on home health claims, patient characteristics, and HHAs from 2008 – the year before the first home health strike force was created. However, we also use home health claims dating back to 2003 for backward-looking characteristics of home health services,

[14] We define full-year-equivalent Traditional Medicare enrollment by counting each enrollee according to the share of months they were enrolled during the year. We use this definition when constructing per-enrollee measures throughout the manuscript.

[15] That is, in a prosecution, the defendant is the HHA, not a particular set of claims "within" the HHA that is suspected of fraud.

[16] We have searched for additional press releases of prosecutions of home health agencies in these districts from 2020 to 2025 and found only three that were for home health agencies that existed in 2008. In other words, our analysis picks up essentially all of the HHAs that were operating in 2008 and subsequently prosecuted for fraud.

such as changes in doctor prescription behavior within an HHA from year to year. Our maintained assumption is that 2008 has yet been affected by the threat of fraud prosecution, making it a clean, "non-contaminated" sample in which HHAs have yet to adjust their behavior to the threat of prosecution. We then use the subsequent prosecutions as revealing HHAs' true type – fraudulent or not – for those that were operating in strike force districts in 2008. We construct HHA-level features to use as predictors, and apply a standard Machine-Learning technique to predict HHA type.

Feature construction. We begin by linking each 2008 home health claim to its associated patient episode, billing HHA, and prescribing doctor. This yields roughly 32.4 million claims for our final 2008 sample.[17] For each claim, we merge on myriad characteristics of its associated parties, such as patient demographics and chronic conditions, contemporaneous and lagged doctor prescription behavior, and HHA ownership structure and available services. We intentionally do not include any location information, such as provider or patient ZIP code, as we will later use the predictions out of sample, in districts that did not receive a strike force.

Since prosecution occurs at the HHA level (rather than claim level), we aggregate claim characteristics within each HHA to generate HHA-level potential predictors. For example, for various claim-level continuous variables (e.g., number of visits, episode length, or amount billed), we compute, for each HHA, first, second, and third moments of their distribution within the HHA, as well as their 5th, 50th, and 95th percentiles. For categorical variables (e.g., type of services), we compute modes and concentration measures. In total, we construct more than 300 potential, HHA-level predictors. A full list of predictors is available in Appendix Tables OA.6a through OA.6d.

Broadly speaking, the potential predictors capture four aspects of each HHA. One concerns information on HHA characteristics, such as whether it is for-profit or non-profit or when it began operating. A second aspect uses the claim-level data to summarize features of the patient episodes associated with the HHA, such as properties of the distribution of episode payments, duration and timing (e.g. the average number of days between episodes for patients at the HHA). A third type of predictor characterizes the patients associated with the HHA (e.g. their age, race, gender, and comorbidities, as well as their history of Medicaid (i.e. dual-enrollee) coverage and Medicare Advantage coverage; during enrollee-months when the individual is covered by Medicare Advantage, we do not observe home health claims or co-morbidities). Finally, we construct predictors that capture the nature of the physicians who certify the episodes associated with each HHA and their patient volume. Overall, we associate each HHA with 55 HHA-related predictors, 56 episode-related predictors, 118 patient-related predictors, and 89 doctor-related predictors.

[17] Approximately 9 million of these claims cannot be linked to a prescribing doctor. We keep these claims in the sample, but omit them when computing summary statistics of doctor variables at the HHA level.

Implementation. We start by splitting the nine districts that had a strike force into a training sample and a test sample by randomly assigning 6 districts to training and 3 districts to test. Randomization was performed at the district level to avoid any potential spillovers or spatial correlation contaminating the exercise. Randomization was stratified by the number of home health episodes in 2008 in each district; one district in each tercile was randomly selected for testing, with the other two assigned to training. The result was 2,519 HHAs in the training sample, out of which 77 (3.1%) were prosecuted, and 1,084 HHAs in the test sample, out of which 41 (3.8%) were prosecuted. The prosecuted HHAs account for 5.0% of Medicare home health spending in the training sample, and 3.5% in the test sample. We fit the model on the training sample, and evaluate its out-of-sample performance on the test sample.

Because only 3.1% of the HHAs in the training sample were prosecuted, the sample is highly imbalanced. We thus adopt the framework of "imbalanced learning" and use the Imbalanced Learn package's Balanced Random Forest classifier which is a random forest algorithm that has been customized for prediction in imbalanced datasets (Lemaître et al. 2017). Unlike a standard random forest, it balances the positive and negative classes within each tree's bootstrap sample. Thus, each tree is fit on a balanced sample, but the forest uses the complete sample of data over the course of training across all trees.

We tune the random forest hyper-parameters to minimize the log-loss of the balanced forest's predicted probabilities. That is, when given hyper-parameter space Θ, we choose the vector θ which produces a probability estimate ($\hat{p}_\theta(x_i)$) that HHA i with characteristics x_i is fraudulent according to the following objective:

$$\theta^* = arg \min_{\theta \in \Theta} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left(y_i \log \hat{p}_\theta(x_i) + (1 - y_i) \log(1 - \hat{p}_\theta(x_i)\right) \right\}, \tag{1}$$

where y_i is an indicator variable for whether or not HHA i was prosecuted or not. This objective rewards both *discretion* (correctly ordering HHAs by fraud probability) and *calibration* (estimating meaningful cardinal scores). We select the depth of the trees, the number of features per tree, the minimum leaf size, and minimum samples per split using cross-validation within the training sample to minimize the objective function in equation (1). As in the split of training vs. test samples, we stratify randomization across folds at the district level to avoid any cross-fold contamination. We enforce that each fold contains at least 25 fraud cases to reduce variability due to small within-fold samples.

Given that individual trees of the random forest are balanced, a balanced random forest generally produces accurate orderings of the data, but has a base rate of positive class predictions which is much higher than the fraud rate in the full (imbalanced) data. We thus recalibrate the probabilities by fitting a (logit) function to the predicted probabilities of the training sample against the true labels. We select the optimal function curvature parameter as well as the Ridge (L_2) penalty for the recalibration using the same cross-validation procedure.

Our model scores well on traditional imbalanced learning benchmarks. Models must trade off precision – the fraction of cases we predict to be fraudulent which are actually fraudulent – against recall – the fraction of actually fraudulent cases we predict to be fraudulent. We plot the relationship between these measures in Panel (a) of Appendix Figure OA.1, separately for the six training districts and the three test districts. By integrating under this curve, we compute our model's *average precision* in the test districts of 16.5%. This means that for a randomly selected level of recall, we would expect 16.5% of our predicted positive class to be true positives; this represents 4-fold improvement over the random chance rate of 3.8% (as represented by the dashed horizontal line in the figure).

The Receiver Operating Characteristic (ROC) curves for both test and training samples are shown in Panel (b). The area under this curve (known as the ROC-AUC) in the held-out test sample is 74.7%; in other words, if one randomly selected an actual positive and actual negative HHA, the model would correctly sort the examples 74.7% of the time.

5 Results

5.1 Fraud amount and its geographic distribution

Our estimates indicate that home health fraud represents a modest but non-negligible share of total Medicare home health spending. Across all strike force districts, 4.5% of home health spending was billed by HHAs that were prosecuted for fraud. In the held-out test sample, 3.5% of home health dollars were predicted as fraudulent,[18] exactly matching the true rate of 3.5%, suggesting good model fit. Appendix Figure OA.2 shows the full calibration curve.[19] Extrapolating to untreated districts, we predict that approximately 2.0% of total home health spending in those areas is accounted for by fraudulent HHAs. In total, across all the districts, we estimate that 3.4% of home health care spending – approximately $520 million – was associated with fraudulent HHAs in 2008.[20]

The strike forces appear to be well targeted. Although the nine home health strike force districts that opened between 2009 and 2013 accounted for only 22.6% of Medicare enrollees and 39% of Medicare home health spending in 2008, we estimate that they contained 64.9% of home health spending associated with fraudulent HHAs. Moreover, the remaining fraudulent activity

[18] As noted in Section 4, the training sample had a higher base rate of home health spending in prosecuted HHAs of 5.0%, which can happen by chance.

[19] Our model exhibits significant under-confidence within the training data. This is expected of random forests applied to their own training data when in-sample discretion is high (Barreñada et al. 2024).

[20] This reflects the amount paid by Medicare, which is generally less than the amount billed to Medicare. For reference, estimates based on our analysis of amounts disclosed in DOJ press releases show that the 118 prosecuted HHAs in our sample submitted approximately $2.2 billion in fraudulent claims, of which about $1 billion was ultimately paid by Medicare.

was unevenly distributed across the country. Figure 2 reports the predicted home health fraud dollars per full-year equivalent Traditional Medicare enrollee by district in logs. Outside of the strike force districts, we find particularly high predicted rates of home health fraud per enrollee in several Southern states, including Texas, Louisiana, Mississippi, and Tennessee, and also in Utah.

Of course, since strike forces involve a combination of fixed and variable costs, we would expect strike forces to be targeted at districts that have large amounts of home health fraud, not just large amounts of home health fraud per capita. To investigate this, Appendix Table OA.3 reports home health dollars at fraudulent HHAs per full-year equivalent Traditional Medicare enrollee, and Appendix Table OA.4 reports the total Medicare home health dollars at fraudulent HHAs. For each we report the top twenty districts by each measure, as well as any districts which received strike forces if they are not in the top 20. The strike force districts are clustered at the top of the distribution by both measures. The Southern District of Florida, which includes Miami, appears to be a particular outlier, with home health spending per enrollee at fraudulent HHAs significantly above the level of the next highest district by either measure.

5.2 Predictors of fraudulent HHAs

The balanced random forest identifies a set of systematic differences between prosecuted and non-prosecuted HHAs. These differences fall into several domains. To illustrate, Figure 3 summarizes the impact of selected strong predictors from among the top 30 predictors with the highest variable importance.[21] It shows that fraudulent HHAs tend to have a lot of their episodes prescribed by high-prescribing physicians, operate with extremely uniform and low-intensity episode structures, and serve unusually young and healthy beneficiaries.

Super-Prescribers. Fraudulent HHAs disproportionately rely on extremely high-volume referring physicians. Panel (a) of Figure 3 shows that the fraction of episodes at an HHA coming from doctors with more than 10,000 annual HHA visits in 2008 – a level far outside normal practice patterns – is a strong predictor of fraud. Figure OA.3 shows related evidence: fraudulent HHAs tend to have a higher share of their episodes prescribed by referring physicians whose referrals result in unusually high HHA visits, and by doctors with significantly more prescriptions relative to the previous year.

Episode homogeneity. Fraudulent HHAs display unusually uniform patterns of care. Their episode-level distributions have too little variation and too little skew relative to legitimate providers. Non-fraudulent HHAs show the expected right-skewed distribution of visits and payments – some patients require intensive care, others minimal – whereas fraudulent agencies appear to bill nearly identical episode amounts across patients. Panel (b) of Figure 3 shows this pattern: fraudulent

[21]For completeness, Appendix Table OA.5 lists all of the 30 most important predictors.

HHAs display very little skewness in the number of visits per episode. Figure OA.4 further reports the coefficient of variation in the number of episodes, the standard deviation in the number of visits per episodes, as well as the 95th percentile in visits and payments per episode. Across all measures, episodes at fraudulent HHAs appear remarkably uniform relative to those at other HHAs, a pattern that is consistent with using highly standardized, or "rubber-stamped" episode structures.

Service composition. Fraudulent HHAs differ in service intensity and the types of services they claim. Panel (c) of Figure 3 shows that fraudulent HHAs tend to have lower average payment per episode of care. Figure OA.5 further shows that fraudulent HHAs tend to have very few short-stay episodes that would receive a reduced payment through the low utilization payment adjustment (LUPA) described in Section 2.

Patient characteristics. Fraudulent HHAs tend to serve a distinctive patient mix. Panel (d) of Figure 3 shows that the patients tend to be younger at fraudulent HHAs. Figure OA.6 further shows that fraudulent HHAs have a lower share of patients with hypothyroidism, atrial fibrillation, and hip fractures. This pattern is consistent with prosecutions documenting the recruitment of beneficiaries who are too healthy to plausibly meet Medicare's home-bound eligibility criteria.[22]

6 Conclusion

Concerns about high rates of Medicare fraud – and hopes that eliminating it could save taxpayer dollars without harming patients – have long animated policy discussions. With this dual motivation, we estimated the amount of fraud in Medicare home health, an area considered particularly prone to fraud and the subject of geographically targeted federal strike forces. We used the characteristics of home health agencies that were and were not prosecuted for fraud in nine strike-force districts to train a machine learning classification algorithm to predict the probability that each HHA in non-strike-force districts would have been prosecuted if a strike force had entered.

We estimate that in 2008, prior to the entry of the strike forces, 3.4% of Medicare home health spending – about $520 million – was billed by fraudulent HHAs. The nine districts that had strike forces accounted for only 40% of Medicare home health spending but 65% of spending at fraudulent HHAs.

These results should be interpreted with several caveats. First, the analysis infers fraud probabilities from observed characteristics of HHA prosecutions; unobserved enforcement selection may bias estimates if strike forces targeted certain types of HHAs or providers. Second, we define "fraud" based on what was successfully prosecuted, which in turn reflects current legal standards

[22] A final set of predictors concerns beneficiaries' insurance history. The interpretation of these predictors is less clear. Medicaid status (dual-eligibles) may capture socioeconomic vulnerability or patient recruitment patterns. Medicare Advantage enrollment history may reflect incomplete prior fee-for-service claims.

of evidence and the resources of the strike forces; under different standards of evidence, different resources, or different detection technologies, the amount of prosecuted fraud might well differ. Third, machine-learning importance measures reveal associations rather than causal effects; characteristics of HHAs that are highly predictive of fraud – such as uniform visit patterns across episodes or reliance on extremely high-volume referring physicians – are useful signals for enforcement, but are not direct evidence of intent or wrongdoing.

Despite these caveats, two points are worth emphasizing. First, since Medicare home health is considered particularly susceptible to fraud, the share of fraudulent spending in other parts of Medicare is likely lower than what we estimate for home health. Second, our estimates suggest that while prosecuting Medicare fraud could yield meaningful absolute savings, in proportional terms it would not put much of a dent in the trillion-dollar annual Medicare budget.

References

42 CFR 424.22, "42 CFR § 424.22 Requirements for home health services.," https://www.law.cornell.edu/cfr/text/42/424.22 2020.

Alexander, Diane and Molly Schnell, "The impacts of physician payments on patient access, use, and health," *American Economic Journal: Applied Economics*, 2024, *16* (3), 142–177.

Barreñada, Lasai, Paula Dhiman, Dirk Timmerman, Anne-Laure Boulesteix, and Ben Van Calster, "Understanding Overfitting in Random Forest for Probability Estimation: A Visualization and Simulation Study," 2024, *8*, 14.

Brot-Goldberg, Zarek C, Amitabh Chandra, Benjamin R Handel, and Jonathan T Kolstad, "What does a deductible do? The impact of cost-sharing on health care prices, quantities, and spending dynamics," *The Quarterly Journal of Economics*, 2017, *132* (3), 1261–1318.

_ , Samantha Burn, Timothy Layton, and Boris Vabson, "Rationing medicine through bureaucracy: authorization restrictions in medicare," Technical Report, National Bureau of Economic Research 2023.

Bureau of Labor Statistics, "Labor Force Statistics from the Current Population Survey," https://www.bls.gov/cps/aa2019/cpsaat18.htm 2019.

Cabral, Marika and Mark R Cullen, "The effect of insurance coverage on preventive care," *Economic Inquiry*, 2017, *55* (3), 1452–1467.

Center for Medicare & Medicaid Services, "CMS Imposes First Affordable Care Act Enrollment Moratorium on Chicago-area Home Health Agencies to Combat Fraud and Safeguard Taxpayer Dollars," https://www.cms.gov/newsroom/fact-sheets/cms-imposes-first-affordable-care-act-enrollment-moratorium-chicago-area-home-health-agencies-combat 2013.

_ , "State Operations Manual Chapter 2-The Certification Process," 2020.

_ , "Medicare Benefit Policy Manual," https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/Downloads/bp102c07.pdf 2022.

_ , "Medicare & Home Health Care," https://www.medicare.gov/Pubs/pdf/10969-medicare-and-home-health-care.pdf 2023.

Centers for Medicare & Medicaid Services, "Fraud Prevention System: Second Implementation Year," Report to Congress, U.S. Department of Health & Human Services June 2014. CMS Report to Congress; predictive analytics for Medicare fraud prevention.

Clemens, Jeffrey and Joshua D Gottlieb, "Do physicians' financial incentives affect medical treatment and patient health?," *American Economic Review*, 2014, *104* (4), 1320–1349.

Commission, Medicare Payment Advisory et al., "Healthcare spending and the Medicare program. 2025," 2025.

Cooper, Zack and Fiona Scott Morton, "Out-of-network emergency-physician bills—an unwelcome surprise," *N Engl J Med*, 2016, *375* (20), 1915–1918.

_ , Fiona Scott Morton, and Nathan Shekita, "Surprise! Out-of-network billing for emergency care in the United States," *Journal of Political Economy*, 2020, *128* (9), 3626–3677.

Curto, Vilsa, Liran Einav, Amy Finkelstein, Jonathan Levin, and Jay Bhattacharya, "Health care spending and utilization in public and private Medicare," *American Economic Journal: Applied Economics*, 2019, *11* (2), 302–332.

Cutler, David, "Analysis & commentary how health care reform must bend the cost curve," *Health Affairs*, 2010, *29* (6), 1131–1135.

Department of Health & Human Services & Department of Justice, "Health Care Fraud and Abuse Control Program Annual Report for Fiscal Year 2009," https://oig.hhs.gov/publications/docs/hcfac/hcfacreport2009.pdf 2010.

_ , "Health Care Fraud and Abuse Control Program Annual Report for Fiscal Year 2010," https://oig.hhs.gov/publications/docs/hcfac/hcfacreport2010.pdf 2011.

_ , "Health Care Fraud and Abuse Control Program Annual Report for Fiscal Year 2011," https://oig.hhs.gov/publications/docs/hcfac/hcfacreport2011.pdf 2012.

_ , "Health Care Fraud and Abuse Control Program Annual Report FY 2021," https://oig.hhs.gov/publications/docs/hcfac/FY2021-hcfac.pdf 2022.

Department of Health and Human Services, "Reliance on Unverified Patient Lists Creates a Vulnerability in Home Health Surveys," OEI Report OEI-05-16-00510, U.S. Department of Health and Human Services March 2018. OEI-05-16-00510.

Department of Health and Human Services and Department of Justice, "Health Care Fraud and Abuse Control Program Report, Fiscal Year 2013," Technical Report HCFAC-2013, Office of Inspector General, U.S. Department of Health and Human Services February 2014. Joint annual report required under Section 1128C of the Social Security Act and issued pursuant to the Health Care Fraud and Abuse Control Program.

Doyle, Joseph, John Graves, and Jonathan Gruber, "In hospital care saves money–and lives," *Boston Globe*, 2015.

Einav, Liran, Amy Finkelstein, and Neale Mahoney, "Long-term care hospitals: A case study in waste," *Review of Economics and Statistics*, 2023, *105* (4), 745–765.

_ , _ , Yunan Ji, and Neale Mahoney, "Overlapping Policy Interventions: Evidence from Home Health," Technical Report, National Bureau of Economic Research 2025.

Eliason, Paul J, Riley J League, Jetson Leder-Luis, Ryan C McDevitt, and James W Roberts, "Ambulance Taxis: The Impact of Regulation and Litigation on Health Care Fraud," *NBER Working Paper 29491*, 2021.

Famakinwa, Joyce, "'No Two States Are Alike"': The Home Care Industry's Road To Standardization," https://homehealthcarenews.com/2023/10/no-two-states-are-alike-the-home-care-industrys-road-to-standardization/ 2023.

General Accounting Office, "Medicare and Medicaid: CMS Needs to Fully Align Its Antifraud Efforts with the Fraud Risk Framework," 2017, (GAO-18-88).

Geruso, Michael, Timothy J Layton, and Jacob Wallace, "What difference does a health plan make? evidence from random plan assignment in Medicaid," *American Economic Journal: Applied Economics*, 2023, *15* (3), 341–379.

Gruber, Jonathan, David H Howard, Jetson Leder-Luis, and Theodore L Caputi, "Dying or Lying? For-Profit Hospices and End of Life Care," *NBER Working Paper 31035*, 2023.

Howard, David H and Ian McCarthy, "Deterrence Effects of Antifraud and Abuse Enforcement in Health Care," *Journal of Health Economics*, 2021, *75.*

Kim, Hyunjee and Edward C. Norton, "Effects of the Ten Percent Cap in Medicare Home Health Care on Treatment Intensity and Patient Discharge Status," *Health Services Research*, 2015, *50* (5), 1606–1627.

Lakdawalla, Darius N, "Economics of the pharmaceutical industry," *Journal of Economic Literature*, 2018, *56* (2), 397–449.

Leder-Luis, Jetson, "Can Whistleblowers Root Out Public Expenditure Fraud? Evidence from Medicare," *Review of Economics and Statistics*, forthcoming.

_ **and Anup Malani**, "The Economics of Healthcare Fraud," *NBER Working Paper 33592*, 2025.

Lee, Stephen and Jonathan Skinner, "Reforming Home Health Care Coverage to Reduce Fraud," https://onepercentsteps.com/wp-content/uploads/brief-hh-210208-1700.pdf 2021.

Lemaître, Guillaume, Fernando Nogueira, and Christos K. Aridas, "Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning," *Journal of Machine Learning Research*, 2017, *18* (17), 1–5.

McKnight, Robin, "Home Care Reimbursement, Long-Term Care Utilization, and Health Outcomes," *Journal of Public Economics*, 2006, *90*, 293–323.

Medicare Learning Network, "Home Health Prospective Payment System (HH PPS) Rate Update for Calendar Year (CY) 2019," https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNMattersArticles/downloads/MM10992.pdf 2019.

Medicare Payment Advisory Commission, "Home Health Services," https://www.medpac.gov/wp-content/uploads/import_data/scrape_files/docs/default-source/reports/Mar11_Ch08.pdf 2010.

_ , "Home Health Care Services Payment System," https://www.medpac.gov/wp-content/uploads/2021/10/Mar10_EntireReport.pdf 2017.

— , "A Data Book: Health Care Spending and the Medicare Program," https://www.medpac.gov/document/a-data-book-health-care-spending-and-the-medicare-program-2/ 2022.

National Health Care Anti-Fraud Association, "The Challenge of Health Care Fraud," https://www.nhcaa.org/tools-insights/about-health-care-fraud/the-challenge-of-health-care-fraud/.

National Post-acute and Long-term Care Study, "Biennial Overview of Post-acute and Long-term Care in the United States," https://www.cdc.gov/nchs/npals/webtables/overview.htm#print 2020.

Office of the Inspector General, "Medicare Fraud Strike Force," https://web.archive.org/web/20200901211422/https://oig.hhs.gov/fraud/strike-force/ 2020.

— , "Medicare Fraud Strike Force," https://web.archive.org/web/20260000000000*/https://oig.hhs.gov/fraud/strike-force/ 2022.

O'Malley, A James, Thomas A Bubolz, and Jonathan S Skinner, "The Diffusion of Health Care Fraud: A Bipartite Network Analysis," *Social Science and Medicine*, 2023, *327.*

Orszag, Peter R, "Health Costs Are the Real Deficit Threat," https://www.wsj.com/articles/SB124234365947221489 2009.

Palmetto GBA, "Low Utilization Payment Adjustment," https://palmettogba.com/jmhhh/did/2ji32fdncx~home%20health~home%20health%20patient-driven%20groupings%20model%20(pdgm) 2025. Accessed: 2025-12-31.

Sengupta, Manisha, Jessica P Lendon, Christine Caffrey, Amanuel Melekin, and Priyanka Singh, "Post-acute and long-term care providers and services users in the United States, 2017–2018.," *Vital and Health Statistics*, 2022, *47* (3).

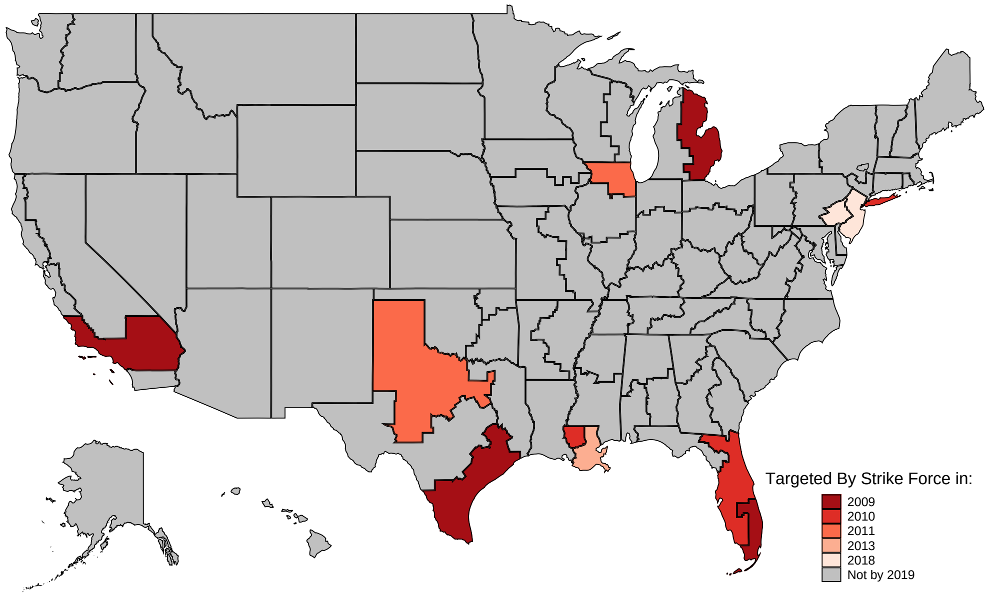
Senior Medicare Patrol, "Dollars Lost to Fraud," https://smpresource.org/medicare-fraud/dollars-lost-to-fraud/ n.d.

Shi, Maggie, "Monitoring for waste: Evidence from medicare audits," *The quarterly journal of economics*, 2024, *139* (2), 993–1049.

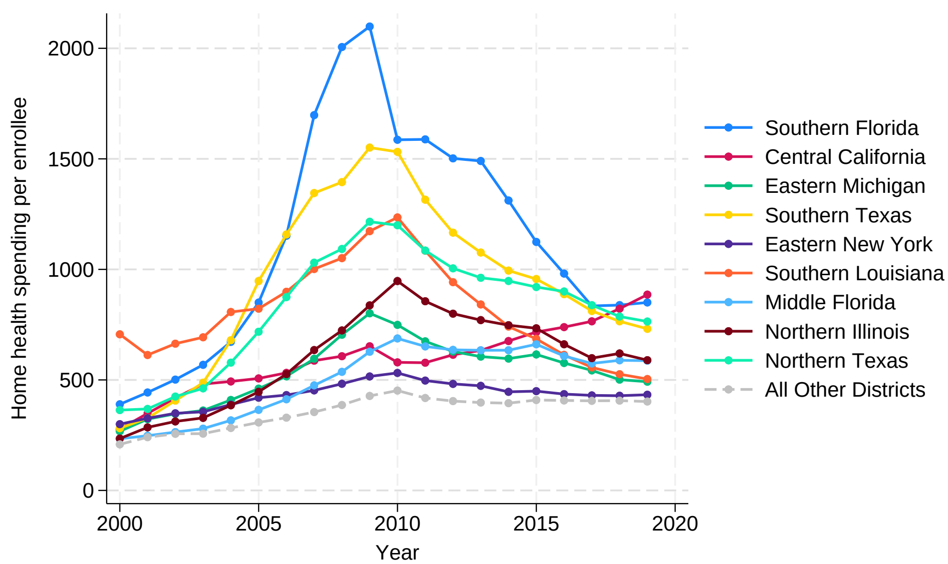
Shrank, William H, Teresa L Rogstad, and Natasha Parekh, "Waste in the US health care system: estimated costs and potential for savings," *Jama*, 2019, *322* (15), 1501–1509.

Figure 1. Characteristics of Strike Force Districts

(a): Geographic Distribution of Strike Forces

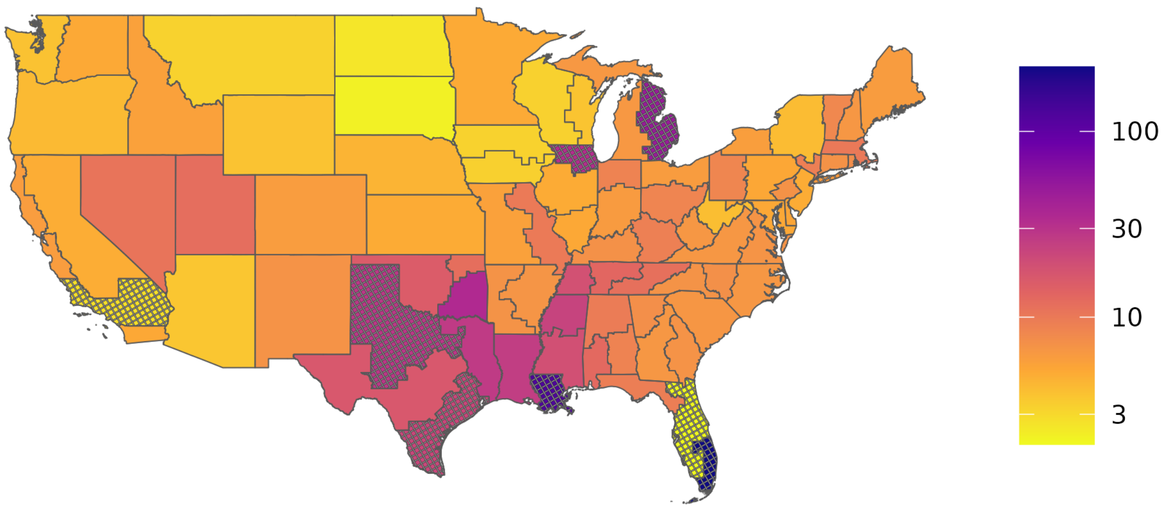


(b): Trends in Home Health Spending per enrollee, 2000-2019



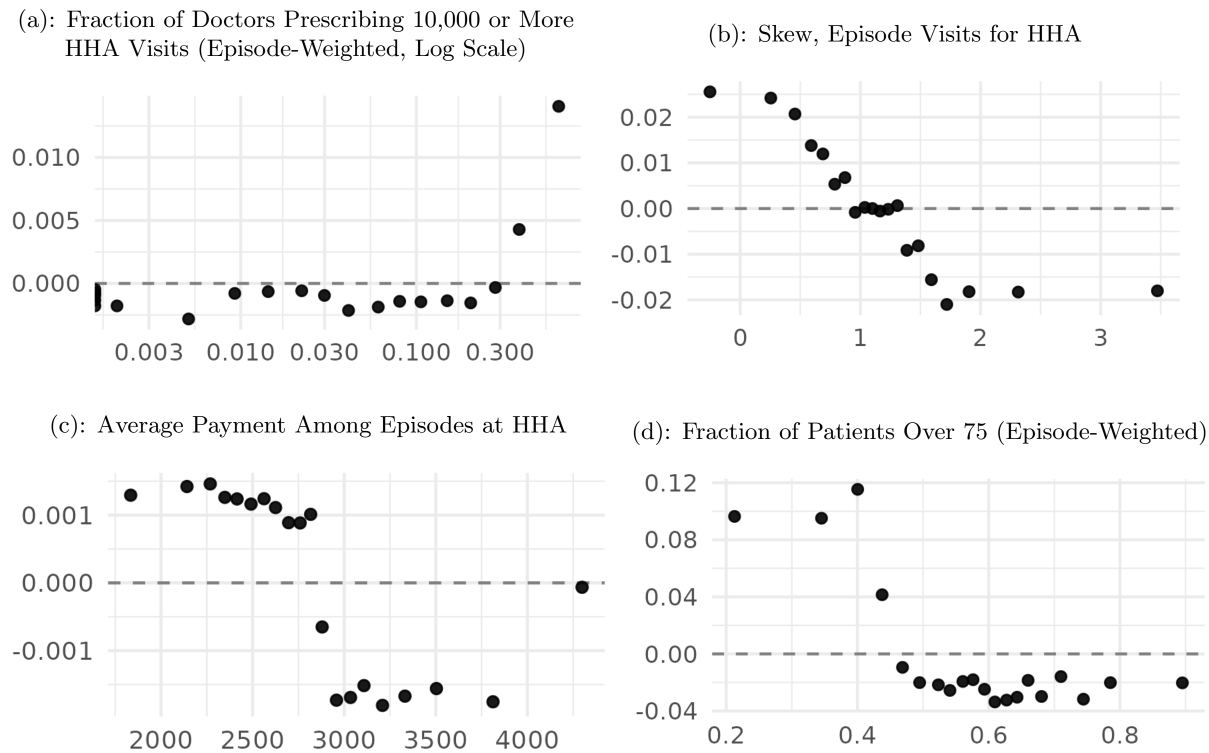
Notes: Panel (a) shows judicial districts in the US by when and whether they were targeted by a home health strike force. Panel (b) reports home health spending per full-year-equivalent Traditional Medicare enrollee from 2000 to 2019 for each of the nine judicial district in the United States that received a strike force prosecuting home health between 2009 and 2013. For districts that did not receive a strike force, we report the enrollee–weighted average home health spending per enrollee, where weights are based on each district's Medicare enrollee population in 2006.

Figure 2. 2008 Per Enrollee Home Health Spending at Fraudulent HHAs, by District (Log Scale)



Note: The map shows a heat map, for each judicial district in the United States, of Medicare home health spending per full-year-equivalent Traditional Medicare enrollee that was billed by fraudulent HHAs in 2008 on a log scale. For districts which received a strike force (marked with a pattern) we sum up all 2008 home health spending associated with fraudulent (i.e. prosecuted) HHAs, then divide by the total number of Medicare enrollees. For districts which did not receive a strike force (no pattern), we compute for each HHA the predicted fraudulent home health spending per Medicare enrollee by multiplying the total home health payments in each HHA in 2008 by our prediction of the probability the HHA would be prosecuted; summing up these predicted fraud dollars across the district, we then divide by the number of Medicare enrollees in the district.

Figure 3. Relationship between HHA-level Observables and Predicted Fraud



Note: Each figure shows the *ceteris paribus* impact of a given observable on fraud probability. For each observable j, we randomly permute the variable (within each district) for each test set HHA observation i, then re-estimate the fraud probability holding all the other variables constant. For each HHA, then, we compare predicted fraud for the truly observed variables $\hat{p}_i$ relative to the permuted benchmark $\tilde{p}_{ij}$ to estimate the "marginal effect" of the variable on predicted fraud. Binning observations by ventile of the "true" observable, each binscatter shows $\ln(\hat{p}_i) - \ln(\tilde{p}_{ij})$, the percentage change in fraud probability for the "true" observed j relative to the permuted baseline. A value of 0 suggests the permutation had no effect; a value above 0 suggests the original value was more indicative of fraud than a randomly selected value, holding all other observables constant. All observables are episode-weighted within the HHA; for example, if doctor A prescribes 10x the cases as doctor B within a given HHA, doctor A's characteristics will receive 10 times as much weight for averages within the HHA.

Appendix: Collecting data on prosecuted HHAs

To identify home health agencies (HHAs) prosecuted for fraud under the strike force, we compiled information from three primary public enforcement data sources: the archived press releases of the Health Care Fraud Prevention and Enforcement Action Team (HEAT), the US Department of Justice (DOJ) website, and the individual U.S. Attorney's Office (USAO) websites of each strike force district. These press releases contain detailed records of indictments, convictions, and sentencing announcements arising from federal health care fraud enforcement activities.

Identifying prosecutions. We first collected information from the HEAT website, which archives enforcement actions of the Medicare Fraud strike force between 2009 and 2014.[23] The website organizes press releases by state and year, covering the strike force districts active by 2014—California, Florida, Illinois, Louisiana, Michigan, New York, and Texas. Each press release was reviewed to determine whether the case involved home health fraud. Articles clearly unrelated to home health services (for example, hospital or physician settlement cases) were excluded, while those potentially related were read in full to verify relevance. For each qualifying case, we recorded the defendant's name, the judicial district where charges were filed, the name of the HHA (if available), the date of the press release, and the enforcement stage (indictment, conviction, or sentencing). URLs for all press releases were retained for documentation and potential follow-up verification.

To extend the coverage of our prosecution data, we supplemented the HEAT archive with additional press releases from the US Department of Justice's official news portal (`https://www.justice.gov/news`). Using the site's search functionality, we entered the keyword "home health" to identify press releases that mention any HHA. The same inclusion criteria and manual review process used for the HEAT dataset were applied to the DOJ press releases to identify additional prosecuted HHAs. Finally, we supplemented the information in the DOJ news portal with additional press releases from individual USAO websites of each strike force district.[24] We again used the same algorithm as we did with the DOJ press releases.

Matching HHAs to Medicare administrative data. Once identified, the names of prosecuted HHAs were matched to their corresponding provider identifiers in Medicare administrative data. We used multiple reference sources for this linkage. First, we matched each HHA's name and location to entries in the Home Health Agency (HHA) directories from 2007, 2009, and 2011. Minor variations in naming conventions (e.g., "Health Care" vs. "Healthcare") were tolerated. When multiple potential matches existed, contextual information—such as the city or district court—was used to select the most plausible HHA.

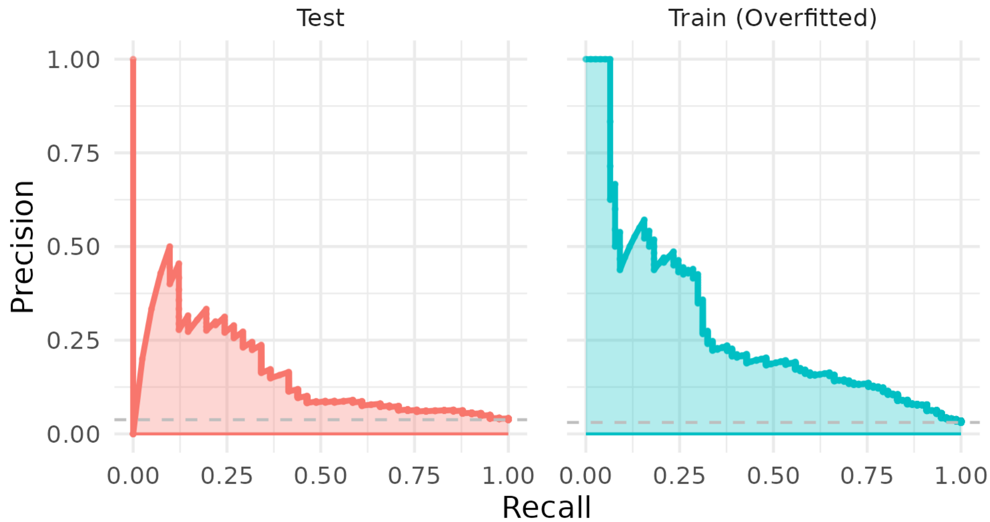
For HHAs not appearing in the directories, we conducted a reverse lookup in the National Provider Identifier (NPI) Registry using the reported agency name and location. When only an NPI was available, we identified the corresponding provider number by cross-referencing the NPI against Medicare claims data. In this way, the vast majority of HHAs could be linked to the unique provider codes used throughout the Medicare claims and Provider of Services (POS) datasets. Our final sample of prosecuted HHAs contains 118 HHAs located in the strike force districts that were in operation in 2008, the year before the first home health strike force.

[23] This link – `https://www.stopmedicarefraud.gov/newsroom/your-state/index.html` – contains press releases through 2014, after which new releases were archived on the DOJ website, `https://www.justice.gov/news`.

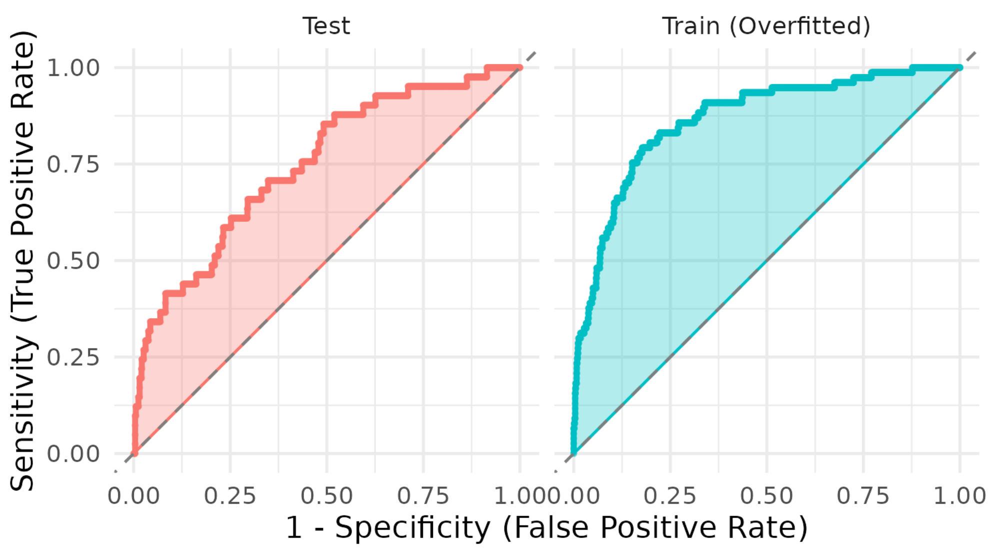
[24] e.g. the press releases for Southern Florida USAO can be found at `https://www.justice.gov/usao-sdfl/pr`

Figure OA.1. Model Discretion

(a): Precision-Recall Curve

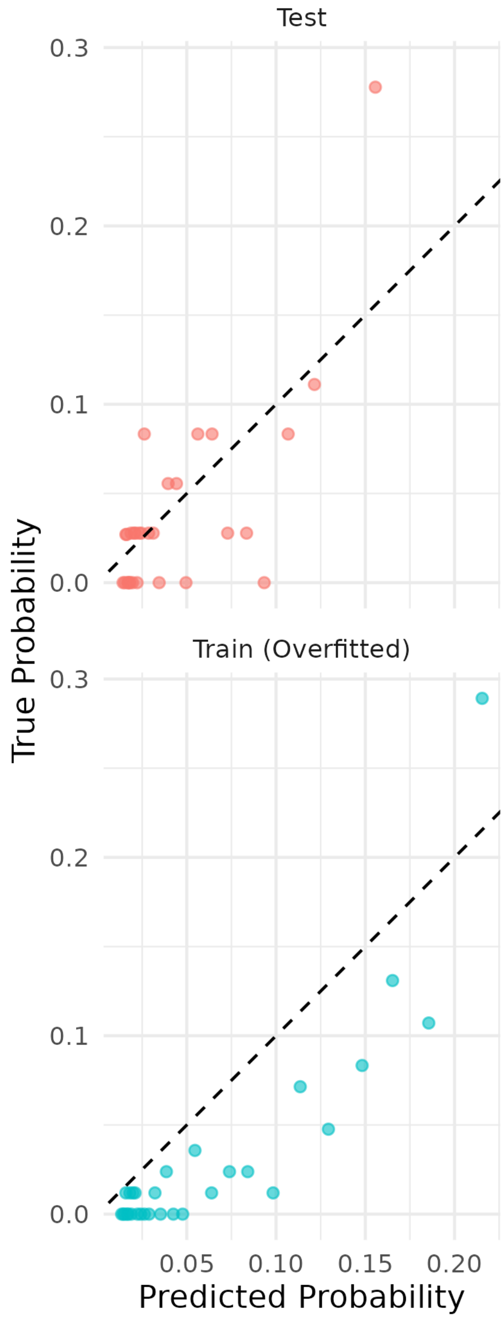


(b): Receiver Operating Characteristic (ROC) Curve



Note: The above figures assess the discretion of our model on both the training data (in blue) and held-out testing data (in red). The model is overfitted to the training data, so the blue figures are overoptimistic about model performance; the red figures constructed from a held-out test set provide a better assessment of the model's performance on fresh data.
Panel (a) shows the precision-recall curve: how the estimator's precision changes as recall increases. The area under this curve is the "average precision" of the model; the larger this area, the better the model. We achieve an average precision of 16.5% in the test sample and 28.8% in the training sample.
Panel (b) shows the receiver operating characteristic curve, plotting the model's true positive rate as we increase the false positive rate. The area under this curve is the "ROC-AUC;" the larger this area, the better the model. We achieve an ROC-AUC of 74.7% in the test sample and 86.5% in the training sample.

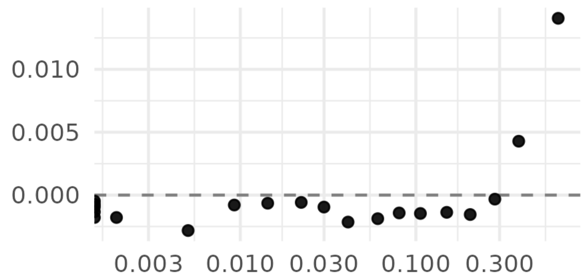
Figure OA.2. Model Calibration



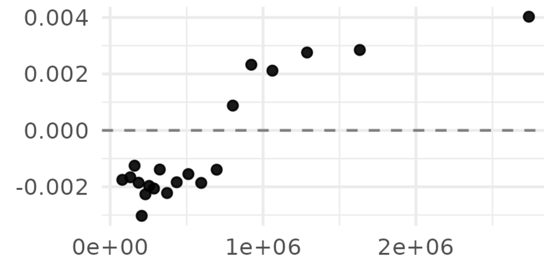
Note: The above figure assesses the calibration of our model on both the training data (in blue) and held-out testing data (in red). The model is overfitted to the training data, so the blue figure is overoptimistic about model performance; the red figures constructed from a held-out test set provide a better assessment of the model's performance on fresh data.
The calibration curve is a binscatter of *true* fraud for each level of *predicted* fraud. That is: for a given predicted probability of fraud $\hat{p}$, what is the true rate of fraud p? The closer the test set is to the 45-degree line (shown), the better the model. Note the miscalibration on the training data is expected (Barreñada et al. 2024).

Figure OA.3. Relationship between HHA-level Observables and Predicted Fraud: Super-Prescribers

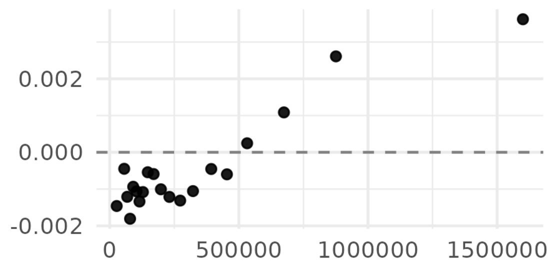
(a): Fraction of Doctors Prescribing 10,000 or More HHA Visits (Episode-Weighted, Log Scale)



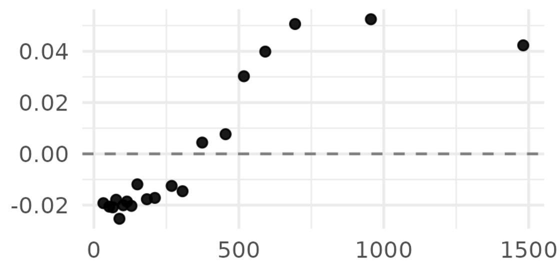
(b): Average Annual Payments Among Doctors at Hospital (Episode-Weighted)



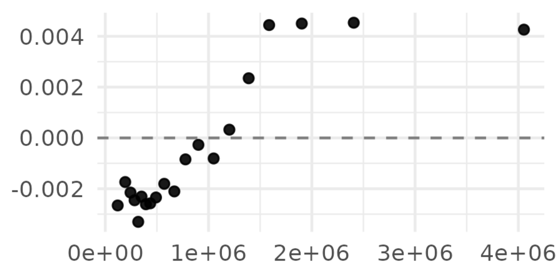
(c): Average Change in Annual HHA Payments per Doctor



(d): Ninety-Fifth Percentile in Changes in Doctor Annual Prescriptions

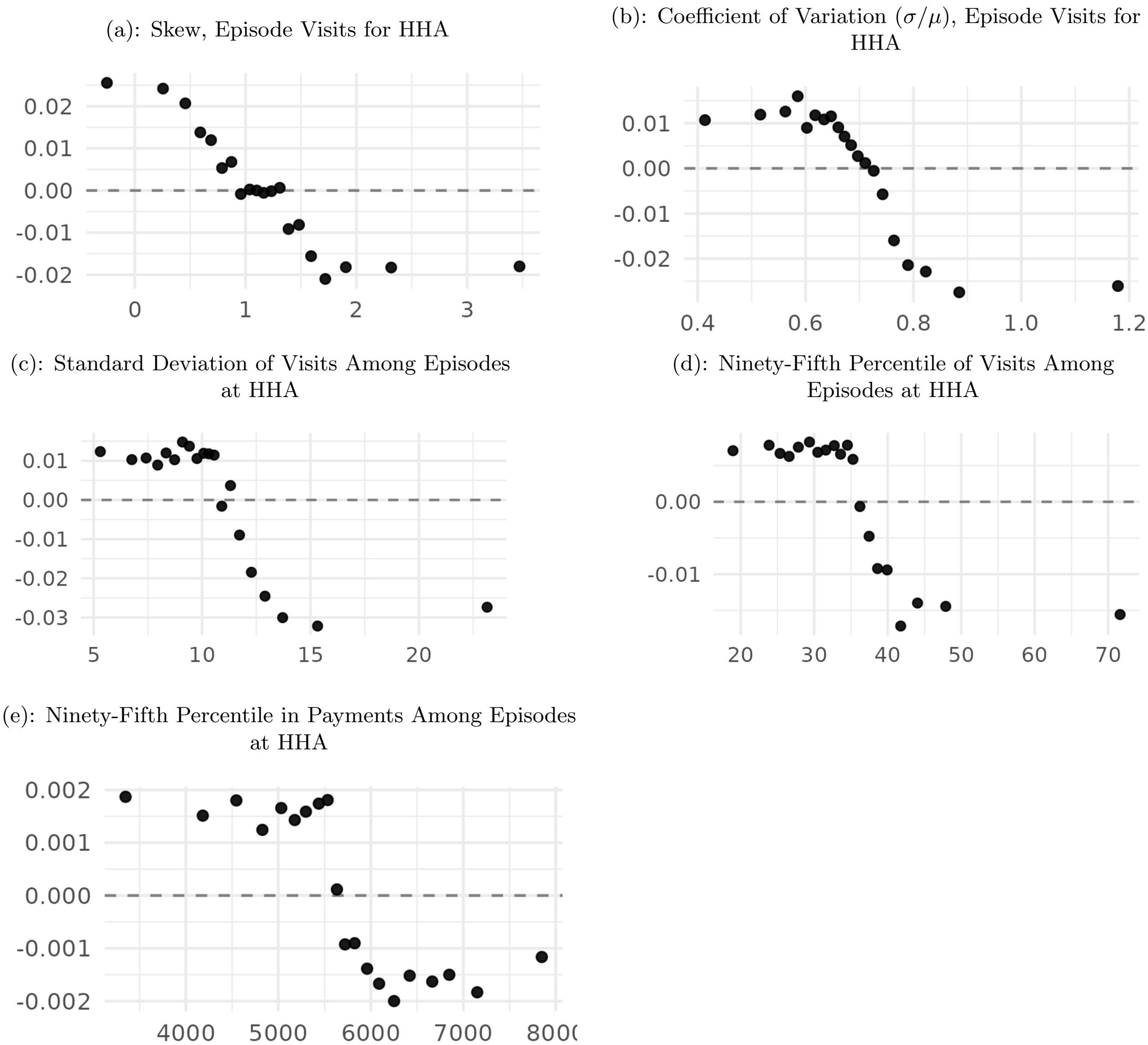


(e): Average Total Payments Among Doctors at HHA (Episode-Weighted)



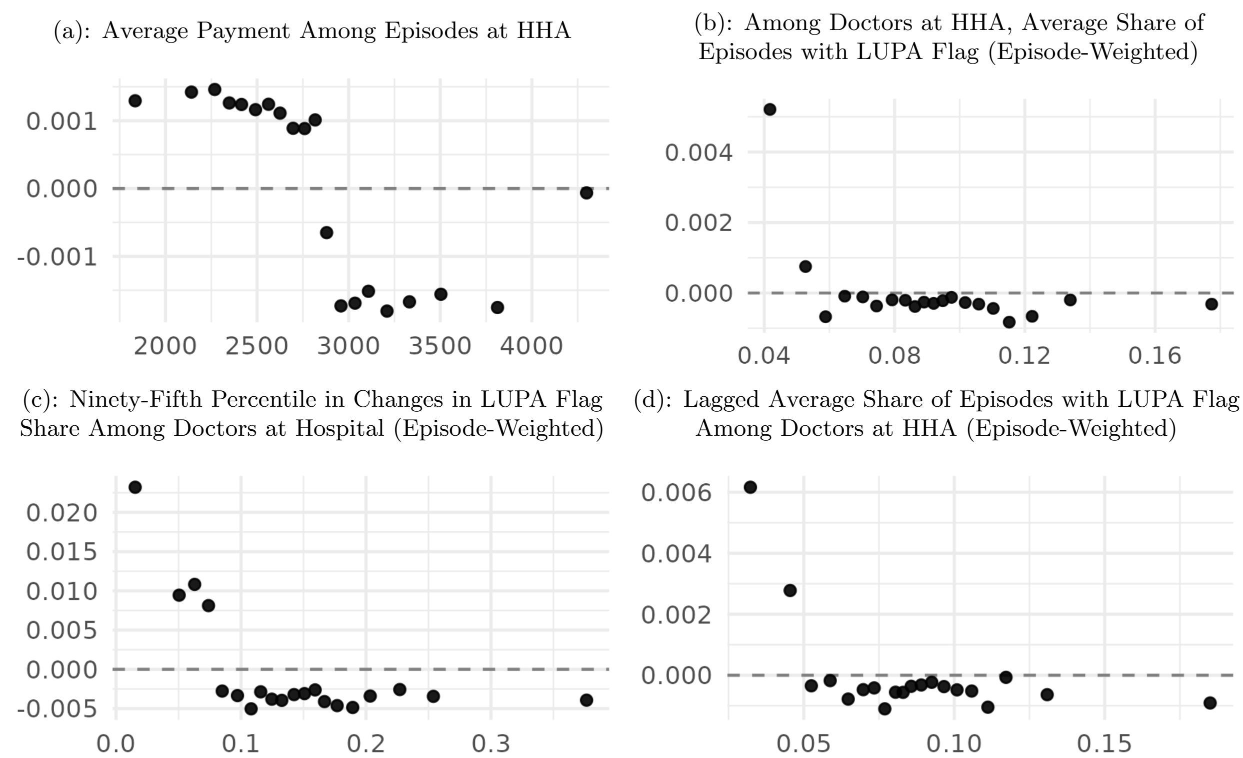
Note: Each figure shows the *ceteris paribus* impact of a given observable on fraud probability. Each observable is an HHA-level average of an episode-level characteristic of the doctor who prescribed that episode. See the note to Figure 3 for more details. Characteristics are based on data from 2008; "change" characteristics use the difference in characteristic from 2007 to 2008.

Figure OA.4. Relationship between HHA-level Observables and Predicted Fraud: Episode Homogeneity



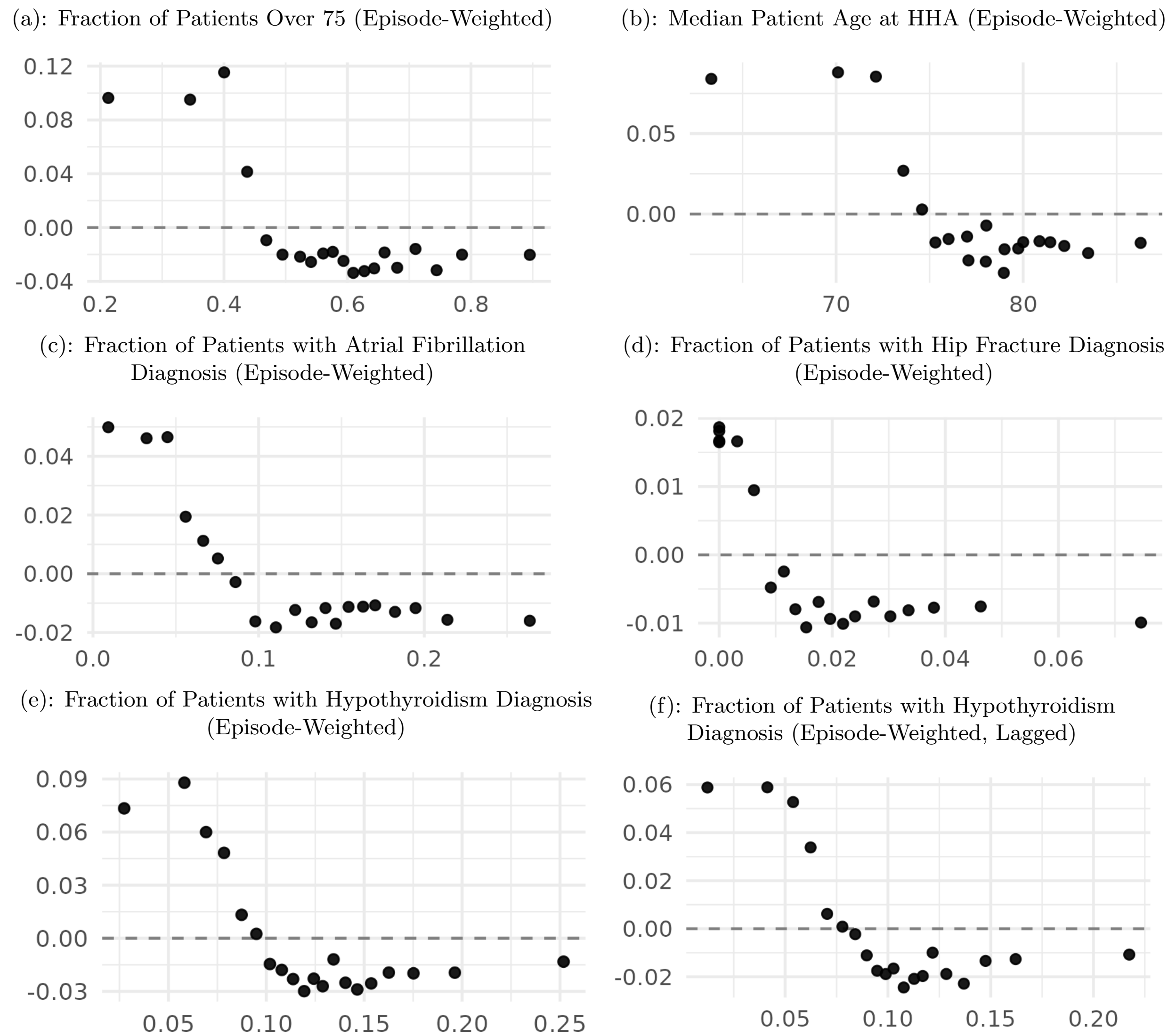
Note: Each figure shows the *ceteris paribus* impact of a given observable on fraud probability. See the note to Figure 3 for more details. Characteristics are based on data from 2008; "lagged" characteristics instead use data from 2007.

Figure OA.5. Relationship between HHA-level Observables and Predicted Fraud: Service Composition



Note: Each figure shows the *ceteris paribus* impact of a given observable on fraud probability. See the note to Figure 3 for more details. Characteristics are based on data from 2008; "lagged" characteristics instead use data from 2007.

Figure OA.6. Relationship between HHA-level Observables and Predicted Fraud: Patient Characteristics



Note: Each figure shows the *ceteris paribus* impact of a given observable on fraud probability. See the note to Figure 3 for more details. Characteristics are based on data from 2008; "lagged" characteristics instead use data from 2007.

Table OA.1. List of Prosecuted HHAs

Investigation Date	State	District	Provider ID	Name
2009-12-03	FL	Southern	108300	Lazaro Home Health Care
2009-12-17	NY	Eastern	337424	Excellent Home Care
2010-08-12	TX	Southern	673132	D'Oro Home Health Services
2010-10-21	TX	Southern	677871	Family Healthcare Group
2011-02-17	IL	Northern	147787	Chalice Home Healthcare Services, Inc.
2011-07-08	CA	Central	058130	Prosperity Home Health Services Inc
2011-11-03	MI	Eastern	237583	Access Care Home Care Inc
2011-11-03	MI	Eastern	237600	Patient Care Home Care Inc
2011-11-03	MI	Eastern	237633	Hands On Healing Home Care Inc
2011-11-03	MI	Eastern	237663	All State Home Care Inc
2011-12-20	FL	Southern	107732	Nany Home Health Inc
2012-01-05	FL	Southern	108147	ABC Home Health Care Inc
2012-01-05	FL	Southern	108298	Florida Home Health Care Providers Inc
2012-01-18	FL	Southern	107796	Priority Home Health
2012-01-26	FL	Southern	108474	Prime Home Health Services Inc
2012-02-28	TX	Northern	677883	Apple of Your Eye Healthcare Services Inc
2012-06-15	CA	Central	058356	GreatCare Home Health, Inc
2012-06-27	IL	Northern	147699	Grand Home Health Care, Inc.
2012-06-27	IL	Northern	147746	Romyst Home Health, Inc.
2012-08-13	IL	Northern	147904	Goodwill Home Healthcare, Inc.
2012-09-25	IL	Northern	147770	Rosner Home Healthcare, Inc.
2012-10-04	TX	Northern	677862	PTM Healthcare Services
2012-10-11	TX	Northern	679225	Ultimate Care Home Health Services, Inc
2012-10-15	FL	Southern	108100	Willsand Home Health Inc
2012-10-26	MI	Eastern	237680	Patient Choice Home Healthcare Inc
2013-02-04	FL	Southern	108408	Serendipity Home Health
2013-03-14	TX	Southern	458361	Houston Compassionate Care
2013-03-19	MI	Eastern	237586	First Michigan Home Health Care, Inc
2013-04-02	MI	Eastern	237486	Acure Home Care Inc
2013-04-17	MI	Eastern	237715	All American Home Care Inc
2013-05-03	MI	Eastern	237572	First Choice Home Health Care Services Inc
2013-05-03	MI	Eastern	237672	Reliance Home Care LLC
2013-05-14	FL	Southern	108338	Ideal Home Health, Inc
2013-07-19	FL	Southern	108336	A Plus Home Health Care, Inc
2013-08-06	MI	Eastern	237417	Guardian Angel Home Health Care
2013-08-15	LA	Southern	197495	South Louisiana Home Health Care Inc
2013-09-05	FL	Southern	108366	Global Nursing Home Health Inc
2013-09-05	FL	Southern	108404	Lovable Home Health Services Corp
2013-09-05	FL	Southern	108440	R&M Health Care Inc
2013-11-21	FL	Southern	108373	Nursemed Home Care Corp

2013-12-12	FL	Southern	108188	Caring Nurse Home Health Care Corp
2014-01-01	LA	Southern	197589	Charter Home Health
2014-04-03	LA	Southern	197213	Health Care Options, Inc.
2014-05-02	FL	Southern	108328	Alephzayn Health Services Inc
2014-05-12	TX	Southern	673156	Allied Covenant Home Health Inc
2014-06-20	TX	Southern	677974	Jackson Home Healthcare, Inc
2014-09-25	LA	Southern	197477	Lexmark Health Care LLC
2014-09-25	LA	Southern	197768	Interlink Health Care Services Inc
2014-09-25	LA	Southern	197772	Memorial Home Health Inc
2014-09-25	LA	Southern	197781	Lakeland Health Care Services Inc
2014-10-01	MI	Eastern	237576	Associates in Home Care Inc
2014-10-01	MI	Eastern	237643	Swift Home Care LLC
2014-10-22	MI	Eastern	237706	Prestige Home Health Services Inc
2014-10-22	MI	Eastern	237730	Royal Home Health Care Inc
2014-10-29	MI	Eastern	237636	Physicians Choice Home Health Care LLC
2014-11-05	FL	Southern	108042	LTC Professional Consultants Inc
2014-11-05	FL	Southern	108475	Professional Home Care Solutions Inc
2014-11-13	FL	Southern	108277	Nation's Best Care Home Health Corp
2015-01-09	FL	Southern	108187	AA Advanced Care Inc
2015-01-14	MI	Eastern	237733	AMB Healthcare Inc
2015-02-06	MI	Eastern	237753	Cherish Home Health Services Inc
2015-02-17	FL	Southern	108360	Longcare Home Health Corporation
2015-03-09	FL	Southern	107619	Recovery Home Care Inc
2015-03-12	LA	Southern	197431	PCAH, Inc. aka Priority Care at Home, Inc. (PCAH) d/b/a Abide Home Care Services Inc.
2015-05-28	FL	Southern	108123	Unlimited Care of Florida Inc
2015-05-28	FL	Southern	109017	USA Home Care Solution Agency Corp
2015-06-02	FL	Southern	109042	Limited Home Health Care Inc
2015-06-04	FL	Southern	109067	Renovation Health Care Inc
2015-06-10	TX	Northern	679388	Essence Home Health
2015-06-11	FL	Southern	108409	D&D&D Home Health Care Inc
2015-06-11	MI	Eastern	237407	At Home Network Inc
2015-06-12	LA	Southern	197231	Christian Home Health Care, Inc
2015-06-16	LA	Southern	197474	Bayou River Health Systems Inc
2015-06-17	FL	Southern	108062	JEM Home Health Care LLC
2015-06-17	MI	Eastern	237544	Professional Home Health Care, Inc
2015-06-17	MI	Eastern	237688	Blue Water Home Health Care, Inc
2015-07-15	TX	Southern	679368	Bona Care
2015-08-27	IL	Northern	147771	HCN Home Healthcare Inc.
2015-12-16	MI	Eastern	237690	Advance Home Health Care Services
2016-04-11	IL	Northern	147877	Donnarich Home Health Care Inc
2016-04-11	MI	Eastern	237682	TriCounty Home Care Services Inc
2016-05-31	FL	Southern	108231	America Home Health Inc
2016-06-02	FL	Southern	108347	Maya Home Health Care Inc
2016-06-02	IL	Northern	147869	TLC Healthcare Services of Illinois

2016-06-07	TX	Northern	679091	Elder Care Home Health Services LLC
2016-06-14	FL	Southern	108418	Golden Home Health Care Inc
2016-06-14	FL	Southern	108490	MA Home Health Agency Inc
2016-06-14	FL	Southern	109001	Nova Home Health Care Inc
2016-06-15	TX	Northern	453169	Timely Home Health Services Inc
2016-06-16	TX	Southern	677906	Preferred Health Services Inc
2016-06-16	TX	Southern	679562	Baptist Home Care Providers Inc
2016-06-16	TX	Southern	679653	Criseven Health Management Corporation
2016-06-16	TX	Southern	679714	Colony Health Services Inc
2016-06-16	IL	Northern	147792	Axis Health Care Services, Inc
2016-06-20	TX	Southern	679624	Niron Health Services
2016-11-11	TX	Southern	679188	Fiango Home Healthcare Inc
2016-12-14	LA	Southern	197597	Comprehensive Nursing and Home Health Service, Inc.
2017-07-13	FL	Southern	108160	Better Care Home Health Services, Inc.
2017-07-13	FL	Southern	108441	Sweet Home Services, Inc.
2017-07-13	IL	Northern	147810	Sure Care Home Health Corp
2017-10-17	IL	Northern	147731	Gateway Health Systems Inc.
2018-05-17	MI	Eastern	237503	Exceptional Home Health Care Inc
2018-06-28	FL	Southern	108234	New Life Home HealthCare Inc.
2018-06-29	FL	Southern	108041	F&E Home Health Care, Inc.
2018-08-23	CA	Central	557743	Star Home Health Resources
2018-09-27	MI	Eastern	237701	United Home Health Care
2018-10-04	TX	Southern	677844	Texas Tender Care
2018-12-17	FL	Southern	108264	Sunshine Home Health Services Inc
2019-01-28	FL	Southern	108411	Humanity Home Health
2019-02-27	FL	Southern	107748	T.L.C. Health Services
2019-03-28	IL	Northern	147995	Hexagram Home Health Care
2019-05-16	LA	Southern	197327	Progressive Home Health
2019-05-20	FL	Southern	108317	Medsel Home Health Care Corp
2019-05-24	FL	Middle	108480	Doctor's Choice
2019-05-31	IL	Northern	147636	Renaissance Home Health Services Inc
2019-09-25	FL	Southern	108186	TC Home Health Care Inc
2019-09-27	MI	Eastern	237611	Unity Home Health Care LLC
2019-09-27	MI	Eastern	237705	Personal Touch

Table OA.2. Shares of Home Health Claims and Payments at Prosecuted Home Health Agencies

State	District	City Contained	Strike Force Entry	Share of Claims	Share of Payments
Florida	Southern	Miami	January 2009	0.075	0.108
California	Central	Los Angeles	January 2009	0.004	0.005
Michigan	Eastern	Detroit	June 2009	0.053	0.061
Texas	Southern	Houston	July 2009	0.015	0.015
New York	Eastern	New York	December 2009	0.014	0.020
Louisiana	Southern	Baton Rouge	December 2009	0.140	0.121
Florida	Middle	Tampa	December 2009	0.004	0.004
Illinois	Northern	Chicago	February 2011	0.045	0.048
Texas	Northern	Dallas	February 2011	0.021	0.025

Note: This table reports the share of home health claims and payments attributable to prosecuted HHAs out of home health claims and payments within each judicial district that received a home health strike force between 2009 and 2013. Prosecutions are defined as HHAs that were subsequently prosecuted during our 2009–2019 study period. Shares are calculated using 2008 home health claims and payments in each district. As noted in footnote 10, we group the Eastern and Middle Districts of Louisiana as "Southern" due to their close collaboration and use the earlier strike force date (Middle District).

Table OA.3. Top Districts by Per-Medicare Enrollee Predicted Fraud Dollars

Rank	State	District	Fraud dollars per capita
1	**Florida**	**Southern**	**$222.25**
2	**Louisiana**	**Southern**	**$125.19**
3	**Michigan**	**Eastern**	**$49.18**
4	Oklahoma	Eastern	$34.81
5	**Illinois**	**Northern**	**$32.92**
6	**Texas**	**Northern**	**$28.64**
7	Texas	Eastern	$26.89
8	Louisiana	Western	$25.66
9	Mississippi	Northern	$23.23
10	**Texas**	**Southern**	**$21.99**
11	Mississippi	Southern	$18.90
12	Tennessee	Western	$18.66
13	Texas	Western	$16.44
14	Oklahoma	Western	$15.36
15	District of Columbia	Only	$13.85
16	Tennessee	Middle	$13.15
17	Alabama	Southern	$12.57
18	Utah	Only	$11.83
19	Oklahoma	Northern	$11.69
20	Tennessee	Eastern	$11.36
37	Pennsylvania	Eastern	$7.04
53	**New York**	**Eastern**	**$6.18**
70	New Jersey	Only	$4.92
84	**California**	**Central**	**$2.77**
87	**Florida**	**Middle**	**$2.07**

Note: Table shows the top 20 districts by home health dollars at fraudulent HHAs per full-year-equivalent-Traditional Medicare-enrollee along with selected other districts. Districts in bold received strike forces and show the true discovered per-capita dollars of Medicare fraud. Non-bold districts show the predicted per-enrollee dollars of Medicare fraud according to our model. Districts in gray received strike forces in the second wave starting in 2018, so are not considered treated but are included with predicted fraud values for completeness. We use realized fraud rather than model-predicted fraud in the strike force districts; per Barreñada et al. (2024), because these districts were used to train the model the estimates will not be well-calibrated.

Table OA.4. Top Districts by Total Predicted Fraud Dollars

Rank	State	District	Total fraud dollars
1	**Florida**	**Southern**	**$136,020,142**
2	**Michigan**	**Eastern**	**$35,238,523**
3	**Illinois**	**Northern**	**$31,921,546**
4	**Louisiana**	**Southern**	**$22,164,641**
5	**Texas**	**Northern**	**$16,852,272**
6	**Texas**	**Southern**	**$13,235,197**
7	Texas	Eastern	$9,596,964
8	Texas	Western	$8,575,184
9	Massachusetts	Only	$7,196,440
10	Louisiana	Western	$6,568,542
11	New Jersey	Only	$5,162,428
12	Ohio	Southern	$4,708,533
13	**New York**	**Eastern**	**$4,517,749**
14	New York	Southern	$4,482,114
15	Mississippi	Southern	$4,040,984
16	Ohio	Northern	$3,993,085
17	Pennsylvania	Eastern	$3,656,963
18	California	Northern	$3,592,068
19	Georgia	Northern	$3,584,677
20	Virginia	Eastern	$3,582,842
24	**California**	**Central**	**$3,368,271**
35	**Florida**	**Middle**	**$2,780,373**

Note: Table shows the top 20 districts by total home health spending at fraudulent HHAs, along with selected other districts. Districts in bold received strike forces and show the true discovered total dollars of Medicare fraud. Non-bold districts show the predicted total dollars of Medicare fraud according to our model. Districts in gray received strike forces in the second wave starting in 2018 but use predicted fraud values. We use realized fraud rather than model-predicted fraud in the strike force districts; per Barreñada et al. (2024), because these districts were used to train the model the estimates will not be well-calibrated.

Table OA.5. Top feature importances from fraud model

Rank	Feature	Importance
1	Skew, Episode Visits for HHA	0.048
2	Among Doctors at HHA, Average Share of Episodes with LUPA Flag (Episode-Weighted)	0.045
3	Fraction of Patients Who Are White (Episode-Weighted)	0.037
4	Fraction of Patients Over 75 (Episode-Weighted)	0.033
5	Fraction of Doctors Prescribing 10000 or More HHA Visits (Episode-Weighted)	0.025
6	Fraction of Patients with Hypothyroidism Diagnosis (Episode-Weighted)	0.022
7	Ninety-Fifth Percentile in Changes in LUPA Flag Share Among Doctors at Hospital (Episode-Weighted)	0.019
8	Fraction of Patients with Atrial Fibrillation Diagnosis (Episode-Weighted)	0.019
9	Ninety-Fifth Percentile in Payments Among Episodes at HHA	0.019
10	Coefficient of Variation (σ/μ), Episode Visits for HHA	0.018
11	Fraction of Patients with Prior-Year Hypothyroidism Diagnosis (Episode-Weighted)	0.018
12	Ninety-Fifth Percentile of Visits Among Episodes at HHA	0.017
13	Lagged Average Months of Medicare Advantage Coverage Among Patients (Episode-Weighted)	0.017
14	Average Months of Medicaid Coverage Among Patients (Episode-Weighted)	0.017
15	Standard Deviation of Visits Among Episodes at HHA	0.016
16	Ninety-Fifth Percentile in Changes in Doctor Annual Prescriptions (Episode-Weighted)	0.016
17	Average Months of Medicare Advantage Coverage Among Patients (Episode-Weighted)	0.016
18	Fraction of Patients with Hip Fracture Diagnosis (Episode-Weighted)	0.015
19	Average Annual Payments Among Doctors at Hospital (Episode-Weighted)	0.015
20	Average Change in Annual HHA Payments per Doctor (Episode-Weighted)	0.014
21	Median Patient Age at HHA (Episode-Weighted)	0.013
22	Fraction of Patients with Prior-Year COPD Diagnosis (Episode-Weighted)	0.013
23	Standard Deviation of Change in Annual HHA Prescriptions per Doctor (Episode-Weighted)	0.013
24	Standard Deviation of Months of Medicare Advantage Coverage Among Patients (Episode-Weighted)	0.012
25	Average Total Payments Among Doctors at HHA (Episode-Weighted)	0.012
26	Fraction of Patients with COPD Diagnosis (Episode-Weighted)	0.012
27	Average Change in Prescriptions Among Doctors at HHA Since Prior Year (Episode-Weighted)	0.011
28	Average Payment Among Episodes at HHA	0.011
29	Average Lagged Months of Medicaid Coverage Among Patients (Episode-Weighted)	0.011
30	Lagged Average Share of Episodes with LUPA Flag Among Doctors at HHA (Episode-Weighted)	0.011

Note: Table lists the top 30 features by importance. By default characteristics are based on data from 2008. "Lagged" characteristics refer to the same characteristic in 2007; "change" characteristics take the difference between 2008 and 2007.

Table OA.6a. HHA-level variables in HHA-level analysis sample

Description
Count of FTE Registered Nurses Employed at HHA
Count of FTE Registered Pharmacists Employed at HHA
Count of Home Health Aides Employed at HHA
Count of Licensed Practical or Vocational Nurses Employed at HHA
Count of Occupational Therapists Employed at HHA
Count of Other Personnel at HHA
Count of Physical Therapists Employed at HHA
Count of Social Workers Employed at HHA
Count of Speech Pathologists and Audiologists Employed at HHA
Date HHA Originally Approved for Medicare
Date HHA was Surveyed
Day of Month HHA Ends Fiscal Year
Day of Month HHA Originally Approved for Medicare
Day of Month HHA was Surveyed
Did HHA Change Ownership Ever
Did HHA Change Ownership Since Last Survey
Dieticians Employed at HHA
Does HHA Operate Any Branches
Facility Type (e.g. Rehab, Skilled Nursing)
HHA Also Medicare Hospice
HHA Eligibility to Participate in Medicare and/or Medicaid
HHA Has Surety Bond
HHA Medicare Compliance Status
HHA Medicare Participation
HHA Original Participation Year
HHA Qualified to Perform Outpatient Physical and Speech Therapy
HHA Received Override for Number of Beds
HHA Sourcing for Aides (By Staff/Under Arrangement)
HHA Sourcing for Appliances (By Staff/Under Arrangement)
HHA Sourcing for Interns and Residents (By Staff/Under Arrangement)
HHA Sourcing for Lab Services (By Staff/Under Arrangement)
HHA Sourcing for Medical Social Services (By Staff/Under Arrangement)
HHA Sourcing for Nursing (By Staff/Under Arrangement)
HHA Sourcing for Nutritional Guidance (By Staff/Under Arrangement)
HHA Sourcing for Occupational Therapy (By Staff/Under Arrangement)
HHA Sourcing for Other Services (By Staff/Under Arrangement)
HHA Sourcing for Pharmacy Services (By Staff/Under Arrangement)
HHA Sourcing for Physical Therapy (By Staff/Under Arrangement)
HHA Sourcing for Speech Therapy (By Staff/Under Arrangement)
HHA Sourcing of Nutritional Guidance (By Staff/Under Arrangement)
HHA in Compliance with Program of Correction
HHA is Subunit of Other Institution
Herfindahl Index of Revenue Code Shares at HHA
Month HHA Ends Fiscal Year
Month HHA Originally Approved for Medicare
Month HHA was Surveyed
Number of Distinct Revenue Codes Used by HHA
Number of HHA Branches
Number of Subunits at HHA
Ownership Type of HHA (e.g. Private, State)
Purpose of Surveying HHA
Total Number of Employees at HHA
Whether HHA Provides Aide Trainings and Evaluations
Year HHA was Surveyed

Note: HHA-level variables used in prediction model. Characteristics are based on data from 2008.

Table OA.6b. Episode-level variables in HHA-level analysis sample

Description
Average Days Between Consecutive Episodes for a Patient at HHA
Average Episode Duration at HHA (Days)
Average Number of Visits Per Episode at HHA
Average Payment Among Episodes at HHA
Coefficient of Variation (σ/μ), Episode Visits for HHA
Coefficient of Variation of Episode Payments at HHA
Fifth Percentile of Days Between Consecutive Episodes at HHA
Fifth Percentile of Episode Duration at HHA (Days)
Fifth Percentile of Episode Payments at HHA
Fifth Percentile of Visits per Episode at HHA
Fraction of Episodes Associated with First-Time Prescribees
Fraction of Episodes Renewed Too Soon (Within 60 Days) at HHA
Fraction of Episodes That Are Repeat Claims for a Patient at HHA
Fraction of Episodes Where Prescribing Doctor Different from Patient's Previous
Fraction of Episodes at HHA with Revenue Code 0270 (Medical/Surgical Supplies, General)
Fraction of Episodes at HHA with Revenue Code 0272 (Medical/Surgical Supplies, Sterile)
Fraction of Episodes at HHA with Revenue Code 0420 (Physical Therapy, General)
Fraction of Episodes at HHA with Revenue Code 0421 (Physical Therapy, Visit)
Fraction of Episodes at HHA with Revenue Code 0430 (Occupational Therapy, General)
Fraction of Episodes at HHA with Revenue Code 0431 (Occupational Therapy, Visit)
Fraction of Episodes at HHA with Revenue Code 0441 (Speech Therapy, Visit)
Fraction of Episodes at HHA with Revenue Code 0550 (Skilled Nursing, General)
Fraction of Episodes at HHA with Revenue Code 0551 (Skilled Nursing, Visit)
Fraction of Episodes at HHA with Revenue Code 0561 (Medical Social Services, Visit)
Fraction of Episodes at HHA with Revenue Code 0570 (Home Health Aide, General)
Fraction of Episodes at HHA with Revenue Code 0571 (Home Health Aide, Visit)
Fraction of Episodes at HHA with Revenue Code 0623 (Surgical Dressings)
Fraction of Episodes with LUPA Flag at HHA
Fraction of Episodes with Zero Visits at HHA
Gini Coefficient of Episode Payments at HHA
Interquartile Range of Episode Payments at HHA
Maximum Number of Visits in a Single Episode at HHA
Maximum Payment in a Single Episode at HHA
Median Days Between Consecutive Episodes for a Patient at HHA
Median Episode Duration at HHA (Days)
Median Number of Visits per Episode at HHA
Median, Episode Payments for HHA
Most Common Month for Episode End at HHA
Most Common Month for Episode Start at HHA
Most Common Referral Source Code for Episodes at HHA
Ninety-Fifth Percentile in Payments Among Episodes at HHA
Ninety-Fifth Percentile of Days Between Consecutive Episodes at HHA
Ninety-Fifth Percentile of Episode Duration at HHA (Days)
Ninety-Fifth Percentile of Visits Among Episodes at HHA
Share of Calendar Months with Any Episode Activity at HHA
Skew, Episode Payments for HHA
Skew, Episode Visits for HHA
Skewness of Episode Duration at HHA
Standard Deviation of Days Between Consecutive Episodes at HHA
Standard Deviation of Episode Duration at HHA (Days)
Standard Deviation of Episode Payments at HHA
Standard Deviation of Visits Among Episodes at HHA
Total Episode Payments at HHA
Total Number of Episodes at HHA
Total Number of LUPA Flag Episodes at HHA
Total Visits Across Episodes at HHA

Note: Episode-level variables aggregated to the HHA level used in prediction model. An episode-level table contains the raw data, which is then aggregated into the above HHA-level variables used for prediction. Characteristics are based on data from 2008.

Table OA.6c. Patient-level variables in HHA-level analysis sample

Description
Average Lagged Months of Medicaid Coverage Among Patients (Episode-Weighted)
Average Months of Medicaid Coverage Among Patients (Episode-Weighted)
Average Months of Medicare Advantage Coverage Among Patients (Episode-Weighted)
Average Patient Age at HHA (Episode-Weighted)
Fifth Percentile of Lagged Months of Medicaid Coverage Among Patients (Episode-Weighted)
Fifth Percentile of Lagged Months of Medicare Advantage Coverage Among Patients (Episode-Weighted)
Fifth Percentile of Months of Medicaid Coverage Among Patients (Episode-Weighted)
Fifth Percentile of Months of Medicare Advantage Coverage Among Patients (Episode-Weighted)
Fifth Percentile of Patient Age at HHA (Episode-Weighted)
Fraction of Patients Over 65 (Episode-Weighted)
Fraction of Patients Over 75 (Episode-Weighted)
Fraction of Patients Who Are Black (Episode-Weighted)
Fraction of Patients Who Are Female (Episode-Weighted)
Fraction of Patients Who Are White (Episode-Weighted)
Fraction of Patients Who Changed State Between Years (Episode-Weighted)
Fraction of Patients Who Died During Observation (Episode-Weighted)
Fraction of Patients of Other Races (Episode-Weighted)
Fraction of Patients with Acute Myocardial Infarction Diagnosis (Episode-Weighted)
Fraction of Patients with Alzheimer's Disease Diagnosis (Episode-Weighted)
Fraction of Patients with Alzheimer's or Related Dementia Diagnosis (Episode-Weighted)
Fraction of Patients with Anemia Diagnosis (Episode-Weighted)
Fraction of Patients with Asthma Diagnosis (Episode-Weighted)
Fraction of Patients with Atrial Fibrillation Diagnosis (Episode-Weighted)
Fraction of Patients with Benign Prostatic Hyperplasia Diagnosis (Episode-Weighted)
Fraction of Patients with Breast Cancer Diagnosis (Episode-Weighted)
Fraction of Patients with COPD Diagnosis (Episode-Weighted)
Fraction of Patients with Cataract Diagnosis (Episode-Weighted)
Fraction of Patients with Chronic Kidney Disease Diagnosis (Episode-Weighted)
Fraction of Patients with Colorectal Cancer Diagnosis (Episode-Weighted)
Fraction of Patients with Congestive Heart Failure Diagnosis (Episode-Weighted)
Fraction of Patients with Depression Diagnosis (Episode-Weighted)
Fraction of Patients with Diabetes Diagnosis (Episode-Weighted)
Fraction of Patients with Endometrial Cancer Diagnosis (Episode-Weighted)
Fraction of Patients with Glaucoma Diagnosis (Episode-Weighted)
Fraction of Patients with Hip Fracture Diagnosis (Episode-Weighted)
Fraction of Patients with Hyperlipidemia Diagnosis (Episode-Weighted)
Fraction of Patients with Hypertension Diagnosis (Episode-Weighted)
Fraction of Patients with Hypothyroidism Diagnosis (Episode-Weighted)
Fraction of Patients with Ischemic Heart Disease Diagnosis (Episode-Weighted)
Fraction of Patients with Lung Cancer Diagnosis (Episode-Weighted)
Fraction of Patients with New Acute Myocardial Infarction Diagnosis (Episode-Weighted)
Fraction of Patients with New Alzheimer's Disease Diagnosis (Episode-Weighted)
Fraction of Patients with New Alzheimer's or Related Dementia Diagnosis (Episode-Weighted)
Fraction of Patients with New Anemia Diagnosis (Episode-Weighted)
Fraction of Patients with New Asthma Diagnosis (Episode-Weighted)
Fraction of Patients with New Atrial Fibrillation Diagnosis (Episode-Weighted)
Fraction of Patients with New Benign Prostatic Hyperplasia Diagnosis (Episode-Weighted)
Fraction of Patients with New Breast Cancer Diagnosis (Episode-Weighted)
Fraction of Patients with New COPD Diagnosis (Episode-Weighted)
Fraction of Patients with New Cataract Diagnosis (Episode-Weighted)
Fraction of Patients with New Chronic Kidney Disease Diagnosis (Episode-Weighted)
Fraction of Patients with New Colorectal Cancer Diagnosis (Episode-Weighted)
Fraction of Patients with New Congestive Heart Failure Diagnosis (Episode-Weighted)
Fraction of Patients with New Depression Diagnosis (Episode-Weighted)
Fraction of Patients with New Diabetes Diagnosis (Episode-Weighted)
Fraction of Patients with New Endometrial Cancer Diagnosis (Episode-Weighted)
Fraction of Patients with New Glaucoma Diagnosis (Episode-Weighted)
Fraction of Patients with New Hip Fracture Diagnosis (Episode-Weighted)
Fraction of Patients with New Hyperlipidemia Diagnosis (Episode-Weighted)
Fraction of Patients with New Hypertension Diagnosis (Episode-Weighted)
Fraction of Patients with New Hypothyroidism Diagnosis (Episode-Weighted)
Fraction of Patients with New Ischemic Heart Disease Diagnosis (Episode-Weighted)
Fraction of Patients with New Lung Cancer Diagnosis (Episode-Weighted)
Fraction of Patients with New Osteoporosis Diagnosis (Episode-Weighted)

Fraction of Patients with New Prostate Cancer Diagnosis (Episode-Weighted)
Fraction of Patients with New Rheumatoid Arthritis/Osteoarthritis Diagnosis (Episode-Weighted)
Fraction of Patients with New Stroke/TIA Diagnosis (Episode-Weighted)
Fraction of Patients with Osteoporosis Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Acute Myocardial Infarction Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Alzheimer's Disease Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Alzheimer's or Related Dementia Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Anemia Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Asthma Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Atrial Fibrillation Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Benign Prostatic Hyperplasia Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Breast Cancer Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year COPD Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Cataract Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Chronic Kidney Disease Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Colorectal Cancer Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Congestive Heart Failure Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Depression Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Diabetes Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Endometrial Cancer Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Glaucoma Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Hip Fracture Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Hyperlipidemia Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Hypertension Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Hypothyroidism Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Ischemic Heart Disease Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Lung Cancer Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Osteoporosis Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Prostate Cancer Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Rheumatoid Arthritis/Osteoarthritis Diagnosis (Episode-Weighted)
Fraction of Patients with Prior-Year Stroke/TIA Diagnosis (Episode-Weighted)
Fraction of Patients with Prostate Cancer Diagnosis (Episode-Weighted)
Fraction of Patients with Rheumatoid Arthritis/Osteoarthritis Diagnosis (Episode-Weighted)
Fraction of Patients with Stroke/TIA Diagnosis (Episode-Weighted)
Lagged Average Months of Medicare Advantage Coverage Among Patients (Episode-Weighted)
Median Patient Age at HHA (Episode-Weighted)
Most Common Part A Termination Code Among Patients at HHA (Episode-Weighted)
Most Common Part B Termination Code Among Patients at HHA (Episode-Weighted)
Most Common Prior-Year Part A Termination Code Among Patients at HHA (Episode-Weighted)
Most Common Prior-Year Part B Termination Code Among Patients at HHA (Episode-Weighted)
Ninety-Fifth Percentile of Lagged Months of Medicaid Coverage Among Patients (Episode-Weighted)
Ninety-Fifth Percentile of Lagged Months of Medicare Advantage Coverage Among Patients (Episode-Weighted)
Ninety-Fifth Percentile of Months of Medicaid Coverage Among Patients (Episode-Weighted)
Ninety-Fifth Percentile of Months of Medicare Advantage Coverage Among Patients (Episode-Weighted)
Ninety-Fifth Percentile of Patient Age at HHA (Episode-Weighted)
Number of Patient Counties Served by HHA
Number of Patient ZIP Codes Served by HHA
Skewness of Patient Age at HHA (Episode-Weighted)
Standard Deviation of Lagged Months of Medicaid Coverage Among Patients (Episode-Weighted)
Standard Deviation of Lagged Months of Medicare Advantage Coverage Among Patients (Episode-Weighted)
Standard Deviation of Months of Medicaid Coverage Among Patients (Episode-Weighted)
Standard Deviation of Months of Medicare Advantage Coverage Among Patients (Episode-Weighted)
Standard Deviation of Patient Age at HHA (Episode-Weighted)
Total Number of Unique Patients at HHA

Note: Patient-level variables aggregated to the HHA level used in prediction model. A patient-level table contains the raw data, which is then merged on patient id onto episode-level data. The episode-level data is then aggregated into the above HHA-level variables used for prediction. By default characteristics are based on data from 2008. "Lagged" characteristics refer to the same characteristic in 2007; "change" characteristics take the difference between 2008 and 2007.

Table OA.6d. Doctor-level variables in HHA-level analysis sample

Description
Among Doctors at HHA, Average Number of Visits Prescribed (Episode-Weighted)
Among Doctors at HHA, Average Share of Episodes with LUPA Flag (Episode-Weighted)
Average 'Weekly Shock' Associated with Claims at HHA (Episode-Weighted)
Average Annual HHA Prescriptions per Doctor (Episode-Weighted)
Average Annual Payments Among Doctors at Hospital (Episode-Weighted)
Average Censored Start Date for Doctors Associated with HHA (Episode-Weighted)
Average Change in Annual HHA Payments per Doctor (Episode-Weighted)
Average Change in Annual HHA Prescriptions per Doctor (Episode-Weighted)
Average Change in Annual HHA Visits per Doctor (Episode-Weighted)
Average Change in Count of LUPA Flag Episodes per Doctor (Episode-Weighted)
Average Change in Fraction of LUPA Flag Episodes Across Doctors at HHA (Episode-Weighted)
Average Change in Prescriptions Among Doctors at HHA Since Prior Year (Episode-Weighted)
Average Lagged Annual HHA Payments per Doctor (Episode-Weighted)
Average Lagged Annual HHA Prescriptions per Doctor (Episode-Weighted)
Average Lagged Annual HHA Visits per Doctor (Episode-Weighted)
Average Lagged Count of LUPA Flag Episodes per Doctor (Episode-Weighted)
Average Number of Prescriptions per Doctor at HHA (Episode-Weighted)
Average Number of Providers per Doctor (Episode-Weighted)
Average Total Payments Among Doctors at HHA (Episode-Weighted)
Average Year of Doctor's First Prescription (Episode-Weighted)
Fifth Percentile of Annual HHA Payments per Doctor (Episode-Weighted)
Fifth Percentile of Annual HHA Prescriptions per Doctor (Episode-Weighted)
Fifth Percentile of Annual HHA Visits per Doctor (Episode-Weighted)
Fifth Percentile of Annual Visit Prescriptions Among Doctors at HHA (Episode-Weighted)
Fifth Percentile of Change in Annual HHA Payments per Doctor (Episode-Weighted)
Fifth Percentile of Change in Annual HHA Prescriptions per Doctor (Episode-Weighted)
Fifth Percentile of Change in Annual HHA Visits per Doctor (Episode-Weighted)
Fifth Percentile of Change in Count of LUPA Flag Episodes per Doctor (Episode-Weighted)
Fifth Percentile of Change in LUPA Flag Share Across Doctors at HHA (Episode-Weighted)
Fifth Percentile of Lagged Annual HHA Payments per Doctor (Episode-Weighted)
Fifth Percentile of Lagged Annual HHA Prescriptions per Doctor (Episode-Weighted)
Fifth Percentile of Lagged Annual HHA Visits per Doctor (Episode-Weighted)
Fifth Percentile of Lagged Count of LUPA Flag Episodes per Doctor (Episode-Weighted)
Fifth Percentile of Lagged LUPA Flag Share Across Doctors at HHA (Episode-Weighted)
Fifth Percentile of Prescriptions per Doctor at HHA (Episode-Weighted)
Fifth Percentile of Total Payments per Doctor at HHA (Episode-Weighted)
Fifth Percentile of Weekly Shock Across Doctors at HHA
Fifth Percentile of Year of Doctor's First Prescription (Episode-Weighted)
Fraction of Doctors Prescribing $100k or More in Year (Episode-Weighted)
Fraction of Doctors Prescribing 100 or More HHA Visits in Year (Episode-Weighted)
Fraction of Doctors Prescribing 10000 or More HHA Visits (Episode-Weighted)
Fraction of Doctors Prescribing More Than 1,000 HHA Visits in Year (Episode-Weighted)
Fraction of Doctors with More Than $1,000 in Annual HHA Payments (Episode-Weighted)
Fraction of Doctors with More Than $10,000 in Annual HHA Payments (Episode-Weighted)
Lagged Average Share of Episodes with LUPA Flag Among Doctors at HHA (Episode-Weighted)
Ninety-Fifth Percentile in Changes in Doctor Annual Prescriptions (Episode-Weighted)
Ninety-Fifth Percentile in Changes in LUPA Flag Share Among Doctors at Hospital (Episode-Weighted)
Ninety-Fifth Percentile of Annual HHA Payments per Doctor (Episode-Weighted)
Ninety-Fifth Percentile of Annual HHA Prescriptions per Doctor (Episode-Weighted)
Ninety-Fifth Percentile of Annual HHA Visits per Doctor (Episode-Weighted)
Ninety-Fifth Percentile of Change in Annual HHA Payments per Doctor (Episode-Weighted)
Ninety-Fifth Percentile of Change in Annual HHA Prescriptions per Doctor (Episode-Weighted)
Ninety-Fifth Percentile of Change in Annual HHA Visits per Doctor (Episode-Weighted)
Ninety-Fifth Percentile of Change in Count of LUPA Flag Episodes per Doctor (Episode-Weighted)
Ninety-Fifth Percentile of Change in LUPA Flag Share Across Doctors at HHA (Episode-Weighted)
Ninety-Fifth Percentile of Lagged Annual HHA Payments per Doctor (Episode-Weighted)
Ninety-Fifth Percentile of Lagged Annual HHA Prescriptions per Doctor (Episode-Weighted)
Ninety-Fifth Percentile of Lagged Annual HHA Visits per Doctor (Episode-Weighted)
Ninety-Fifth Percentile of Lagged Count of LUPA Flag Episodes per Doctor (Episode-Weighted)
Ninety-Fifth Percentile of Lagged LUPA Flag Share Across Doctors at HHA (Episode-Weighted)
Ninety-Fifth Percentile of Prescriptions per Doctor at HHA (Episode-Weighted)
Ninety-Fifth Percentile of Total Payments per Doctor at HHA (Episode-Weighted)
Ninety-Fifth Percentile of Weekly Shock Across Doctors at HHA
Ninety-Fifth Percentile of Year of Doctor's First Prescription (Episode-Weighted)

Skewness of Annual HHA Payments per Doctor (Episode-Weighted)
Skewness of Change in Annual HHA Payments per Doctor (Episode-Weighted)
Skewness of Lagged Annual HHA Payments per Doctor (Episode-Weighted)
Skewness of Total Payments per Doctor at HHA (Episode-Weighted)
Skewness of Weekly Shock Across Doctors at HHA
Standard Deviation of Annual HHA Payments per Doctor (Episode-Weighted)
Standard Deviation of Annual HHA Prescriptions per Doctor (Episode-Weighted)
Standard Deviation of Annual HHA Visits per Doctor (Episode-Weighted)
Standard Deviation of Change in Annual HHA Payments per Doctor (Episode-Weighted)
Standard Deviation of Change in Annual HHA Prescriptions per Doctor (Episode-Weighted)
Standard Deviation of Change in Annual HHA Visits per Doctor (Episode-Weighted)
Standard Deviation of Change in Count of LUPA Flag Episodes per Doctor (Episode-Weighted)
Standard Deviation of Change in LUPA Flag Share Across Doctors at HHA (Episode-Weighted)
Standard Deviation of Doctors' LUPA Flag Shares at HHA (Episode-Weighted)
Standard Deviation of Lagged Annual HHA Payments per Doctor (Episode-Weighted)
Standard Deviation of Lagged Annual HHA Prescriptions per Doctor (Episode-Weighted)
Standard Deviation of Lagged Annual HHA Visits per Doctor (Episode-Weighted)
Standard Deviation of Lagged Count of LUPA Flag Episodes per Doctor (Episode-Weighted)
Standard Deviation of Lagged LUPA Flag Share Across Doctors at HHA (Episode-Weighted)
Standard Deviation of Prescriptions per Doctor at HHA (Episode-Weighted)
Standard Deviation of Providers per Doctor (Episode-Weighted)
Standard Deviation of Total Payments per Doctor at HHA (Episode-Weighted)
Standard Deviation of Weekly Shock Across Doctors at HHA
Total Number of Unique Doctors Prescribing to HHA

Note: Doctor-level variables aggregated to the HHA level used in prediction model. A doctor-level table contains the raw data, which is then merged on doctor id onto episode-level data. The episode-level data is then aggregated into the above HHA-level variables used for prediction. By default characteristics are based on data from 2008. "Lagged" characteristics refer to the same characteristic in 2007; "change" characteristics take the difference between 2008 and 2007.