NBER WORKING PAPER SERIES

THE DIFFUSION OF HEALTH CARE FRAUD: A NETWORK ANALYSIS

A. James O'Malley
Thomas A. Bubolz
Jonathan S. Skinner

Working Paper 28560
http://www.nber.org/papers/w28560

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2021

This research was supported by grants P01AG019783 and U01AG046830 from the National Institute on Aging. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w28560.ack

The Diffusion of Health Care Fraud: A Network Analysis
A. James O'Malley, Thomas A. Bubolz, and Jonathan S. Skinner
NBER Working Paper No. 28560
March 2021
JEL No. I1,K42

ABSTRACT

Many studies have examined the diffusion of health care innovations but less is known about the diffusion of health care fraud. In this paper, we consider the diffusion of potentially fraudulent Medicare home health care billing in the United States during 2002-16, with a focus on the 21 hospital referral regions (HRRs) covered by local Department of Justice anti-fraud "strike force" offices. We hypothesize that patient-sharing across home health care agencies provides a mechanism for the rapid diffusion of fraudulent strategies; we measure such activity using a novel bipartite mixture (or BMIX) network index. First, we find a remarkable increase in home health care activity between 2002 and 2009 in some but not all regions; average billing per Medicare enrollees in McAllen TX and Miami increased by $2,127 and $2,422 compared to a $289 increase in other HRRs not targeted by the Department of Justice. Second, we establish that the HRR-level BMIX (but not other network measures) was a strong predictor of above-average home care expenditures across HRRs. Third, within HRRs, agencies sharing more patients with other agencies were predicted to increase spending the following year. Finally, the initial 2002 BMIX index was a strong predictor of subsequent changes in HRR-level home health billing during 2002-9. These results highlight the importance of bipartite network structure in diffusion and in infection models more generally.

A. James O'Malley
The Dartmouth Institute for Health Policy
and Clinical Practice
and Department of Biomedical Data Science
Geisel School of Medicine at Dartmouth
Williamson Translational Research Building, Level 5
1 Medical Center Drive
Lebanon, NH 03756
James.Omalley@dartmouth.edu

Thomas A. Bubolz
The Dartmouth Institute for Health Policy
and Clinical Practice (emeritus)
23 Alice Peck Day Dr.
Lebanon, NH 03766
Thomas.A.Bubolz@dartmouth.edu

Jonathan S. Skinner
Department of Economics
6106 Rockefeller Hall
Dartmouth College
Hanover, NH 03755
and NBER
jonathan.skinner@dartmouth.edu

Supporting code/data is available at http://www.nber.org/data-appendix/w28560

I. Introduction

Since the landmark study by James Coleman and colleagues (1966) of tetracycline, there has been interest in understanding how new medical technologies diffuse, and especially why they appear to exhibit such pronounced geographic patterns. Less well understood, however, is the process by which new fraudulent innovations diffuse over time and across regions.[1] One previous study suggested that Medicare fraud alone could account for 8 percent of total Medicare spending, or $52 billion in 2017 (GAO, 2017). In this paper, we consider the rapid diffusion during the 2000s of a major source of Medicare fraud: Home health care expenditures.

In the aggregate, there was substantial growth in Medicare expenditures for home health care services, with a more than doubling of expenditures over just 7 years -- from $14.9 billion in 2002 to $33.7 billion in 2009 (in 2016 dollars). However, the increase in expenditures was highly concentrated in just a few regions of the U.S. For example, in the Miami Hospital Referral Region (HRR), home health expenditures rose 302% from $802 in 2002 to $3,229 in 2009 per Medicare enrollee (in 2016 dollars). By contrast, in Los Angeles, home health spending barely budged, from $782 in 2002 to $861 in 2009, a 10 percent increase.[2] Largely in response to the rapid growth of home health spending in Miami, the Department of Justice (DOJ) together with the Department of Health and Human Services (HHS) opened a local Southern Florida strike force to prosecute Medicare fraud in 2007; given its success the program was expanded to 8 other locations by fiscal year 2016.

[1] Baker and Faulkner (2006) studied the diffusion of fraud with an interest in how new investment victims are drawn in, rather than the informal sharing of strategies among perpetrators, as we do below. More recently, Nash et al. (2013) considered the role of social networks in attracting new investors to a Ponzi scheme in British Columbia.
[2] Note that the denominator includes all fee-for-service Medicare enrollees, and not just those receiving home health services.

In a pioneering study, Glaeser, Sacerdote, and Scheinkman (1996), hereafter GSS, suggested that the wide regional variation in financial crime was consistent with a model of peer effects, in which fraudulent financial strategies spread rapidly by learning from criminals who live nearby, and the geographical variation in home health care expenditures is clearly consistent with this approach. However, regional variations in health care are also likely associated with other explanations, including regional differences in underlying health (Wolf and Schoomaker, 2019; Chetty et al., 2016), physician beliefs and patient demand (Cutler et al., 2019), peer effects associated with the quality of clinical care (Weng et al., 2019), or variations across regions in social capital and physician professionalism (Skinner, 2011).

To address these concerns, we focus on the network structure of home health care agencies to better understand the remarkable growth of Medicare fraud during the 2000s in just a few regions of the U.S. We build on Becker's (1968) canonical model of criminal behavior to consider explicitly why some regions are so much more likely to invest in fraudulent activities, and more importantly what this implies for the structure of patient-sharing networks most conducive to the rapid diffusion of fraud. As in Sah (1991), agencies may update their objective probabilities of being caught; these, coupled with networks effects, can generate both rapid diffusion as well as scaling back following criminal and civil legal proceedings against agency owners and physicians (Leder-Luis, 2020).

Based on our theoretical model, we develop a new *bipartite* mixture measure, the BMIX index, that captures the idea that a few patients shared across three, four, or more health agencies would more rapidly speed the diffusion of potentially fraudulent billing strategies than a network structure with many patients shared between just two agencies.[3] By contrast, conventional network measures such as density, transitivity, and betweenness-centrality are unipartite measures that do not directly capture the importance of these bipartite relationships. As a measure of "infection," the bipartite BMIX index could also find applications in other analyses of networks, for example in nursing home employee networks associated with the diffusion of COVID-19 outbreaks (Chen et al., 2020).

Briefly, we find using Medicare claims data results that are consistent with GSS at the macro (and micro) level, with a remarkable degree of regional variation across HRRs in home health spending; the coefficient of variation rose from 0.44 in 2002 to 0.69 in 2009, before dropping to 0.49 in 2016. The network BMIX index varied widely across regions and was strongly associated with per-enrollee home health care expenditures; other network measures predicted little variation in home health expenditures. The BMIX index from the initial year of 2002 was highly predictive of the growth (in either dollar or log terms) of home health care spending for the period 2002-09 and was a strong predictor of the subsequent growth in the number of home health agencies, and whether the region would attract a DOJ strike force office. Finally, we found evidence of peer associations *within* HRRs; home health agencies sharing patients with multiple high-spending agencies were more likely to experience higher expenditures in the following year.[4] In sum, our results suggest an important role for market bipartite network structure in the diffusion of fraudulent behavior.

[3] An analogy might be to the pressure or heat (e.g., enthalpy) in the market; low values of the BMIX index correspond to where patients remain with a single home health agency for all their treatment, while larger values are consistent with high-energy jumps to multiple agencies, whether randomly or because of explicit coordination among interlocking home health agencies.

[4] We are limited to measuring peer *associations* rather than causal peer effects given the lack of randomization of network sharing.

II. The Spatial and Temporal Patterns of Medicare Home Health Care Expenditures

Medicare provides home health care benefits for patients who are homebound, require skilled nursing or occupational and physical therapy. Qualified patients receive care by a qualified home health care agency under the direction of a physician who must sign off on treatment plans. Allegations of improper billing for home health services are often brought under the False Claims Act, under the federal anti-kickback provisions, or under civil penalties (Imperato, 2017). Often a "whistleblower" will be involved who provides key evidence regarding the alleged fraud in return for a share of what the government recovers (Leder-Luis, 2020). In this section, we document the close link between high home health care expenditures and fraudulent activity.

Home health fraud. The population-based measures of home health expenditures capture a combination of truly fraudulent activities, "gray area" utilization (or classic supplier-induced demand) unlikely to be prosecuted, and fully legitimate and clinically appropriate use of home health care services. There is ample anecdotal evidence as to the nature of home health care fraud as detailed in many HHS/DOJ Reports:

> In August 2016, the owner and manager of three Miami-area agencies was sentenced to 20 years in prison and ordered to pay $36.4 million in restitution, joint and several. The owner/manager was convicted of conspiracy to commit health care fraud and wire fraud and conspiracy to defraud the United States and pay health care kickbacks for his role in a $57 million Medicare fraud scheme. According to evidence presented at trial, from approximately 2006 to 2013, he and his co-conspirators purported to provide home health services to Medicare beneficiaries, which were not medically necessary and often were never provided. They paid kickbacks to doctors, patient recruiters and staffing groups, which, in exchange, referred beneficiaries to his home health agencies. (HHS/DOJ, 2017, p. 21)

As is clear from this description, there were several dimensions along which billing revenue could be expanded. One was to increase the number of patients by paying kickbacks to "doctors, patient recruiters, and staffing groups." This increases the universe of home health patients, in part by sharing the same patient across multiple home health care agencies. One individual in Houston was convicted of selling a single patient's information to 100 different home health care providers (GAO, 1997, p. 12).[5] Revenue can also be expanded by increasing billing per patient. This can be done in several ways. The first is to simply extend the use of home health care services over several months or more. By 2000, CMS had moved to a home health "bundle" involving 60 days of services, but that bundle could be extended given the compliant physician certified the additional services. In some cases, different agencies were owned or controlled by the same people or family (in the quotation above, the owner controlled 3 separate agencies) but patient sharing across agencies was not always coordinated.

A common approach to enhancing revenue is upcoding (Dafny, 2005; Silverman and Skinner, 2004; Barros and Braun, 2017; Bauder et al., 2017). This may be as simple as "… in-home treatment that was shorter and less complicated than the claims indicated," or include falsely asserting that patients are homebound, a key requirement for home health care eligibility, or both.[6] Another example of upcoding is when patients are falsely coded as having diabetes and unable to self-administer insulin; this allows the home health agency to bill Medicare for nursing visits to provide insulin shots, even if the visits never occurred.[7] In McAllen TX and Miami

---

[5] Insights can also be gained from a 2003 pamphlet for Medicare and Medicaid enrollees (CMS, 2003, p. 21-22) where enrollees are encouraged to look for: "home health visits that your doctor orders that you never get; visits by home health staff that are not needed; bills for services and equipment you never get; faking your signature or your doctor's signature; pressure to accept items and services that you don't need; and items listed on your Medicare Summary Notice that you don't think you received….. You also should be careful about activities such as [h]ome health services your doctor didn't order…. [and] a home health agency that offers you free goods or services in exchange for your Medicare number.

[6] https://www.justice.gov/usao-ndil/pr/federal-jury-convicts-tinley-park-physician-medicare-fraud-scheme

[7] For example, AARP warned its members about a scam involving calls to Medicare enrollees with diabetes: https://www.aarp.org/money/scams-fraud/info-06-2010/scam_alert_fraudsters_target_people_with_diabetes_.html

during 2009, for example, we found nearly two-thirds of Medicare home health patients were coded as having diabetes, in contrast to roughly one-third in the rest of the country.

Geographic variation in home health care expenditures 2002-16. Home health care expenditures per Medicare fee-for-service enrollee are derived from age-sex-race-adjusted measures in the Dartmouth Atlas from 2002-16.[8] The comparisons hold prices constant across regions using constant-price methods documented in Gottlieb et al. (2010). The measure therefore captures both the number (and reimbursement rate) of services per patient receiving home health care, and the fraction of the population receiving any services. All expenditure measures further adjust for (within-year) differences in age, sex, and race across regions, and are adjusted for inflation using the GDP deflator, expressed in 2016 dollars.[9]

There are 306 regions in the U.S., each likely experiencing some degree of fraudulent behavior. However, we consider in detail regions in which by 2016 the Department of Justice had located special strike forces on health care fraud. Following the first office opened in Southern Florida in 2007, by 2016, the DOJ had a total of 9 offices: "Los Angeles, California; Miami and Tampa, Florida; Chicago, Illinois; Brooklyn, New York; Detroit, Michigan; Southern Louisiana; and Dallas and Southern Texas." (HHS/DOJ, 2017; p. 10) Based on this description, along with a 2020 documentation of strike force activity that referred to a "Gulf Coast" office,[10]

[8] https://www.dartmouthatlas.org/

[9] The spending measures are normalized to the age, sex, and race distribution during the specific year, but are not normalized to a common age distribution over time. However, the change in the distribution of age conditional on being 65+ evolves slowly.

[10] Since 2016, additional regions have been added to address the opioid epidemic. As of 2020, the strike force locations are listed as "Florida (Miami, Tampa, Orlando), Los Angeles, CA, Texas (Houston, Dallas, McAllen/Rio Grande), The Gulf Coast (New Orleans, Baton Rouge, and Southern Mississippi), Detroit, MI, Brooklyn, NY, Chicago, IL, Newark, NJ and Philadelphia, PA, Washington, DC (National Rapid Response Strike Force), Kentucky, Ohio, Virginia, and West Virginia (ARPO North), Tennessee and Northern Alabama (ARPO South)" (https://www.justice.gov/criminal-fraud/health-care-fraud-unit, accessed December 16, 2020).

we designated 21 regions deemed subject to strike force interest.[11] To provide graphical clarity in our graphs, we focus on 9 of the larger regions.

Figure 1 shows the time-series of these 9 regions, plus a population-weighted average of the 285 HRRs not included in the geographical districts targeted by the DOJ; these are listed as "Other HRRs." While our formal network analysis begins in 2002 when the 100% fee-for-service data became available, we show in Figure 1 the Dartmouth Atlas data beginning in 2000 (with 20% samples) and running through 2016 to demonstrate that 2002 appeared to be an inflection point.[12] The first thing to note is that for "other" HRRs not explicitly targeted by the DOJ rates were generally low, although there was an increase from $404 in 2002 to $693 in 2009; a large proportional increase (71%) but in dollar terms per enrollee ($289) a barely perceptible change relative to targeted regions.

Second, the Department of Justice location of their local strike force offices were largely (but not exclusively) associated with very high rates of home health care expenditures. McAllen and Miami were roughly 6-times the average rates of the other HRRs, while Chicago, Dallas, and New Orleans, were roughly three times the rate; Detroit and Tampa were double. The offices in Brooklyn and Los Angeles were presumably focused on other types of fraud; the Manhattan HRR (which includes Brooklyn) tracked the non-targeted average closely. Another way to view the predictive value of the DOJ field offices is to note that among the 15 HRRs with the highest level of home health spending in either 2009 or 2010, 11 of them are in our designated targeted

[11] The regions included Gulf Coast HRRs from McAllen to Gulfport. The 21 HRRs are Miami, Manhattan (includes Brooklyn), Chicago, Tampa, Detroit, Los Angeles, Dallas, Corpus Christi TX, McAllen TX, Harlingen TX, Victoria TX, Houston, Beaumont TX, Lake Charles TX, Houma LA, Lafayette LA, New Orleans, Baton Rouge LA, Metairie LA, Slidell LA, and Gulfport, MS.

[12] There is a shift in 2003 from one data series, the Continuous Medicare History Sample File, to the individual 20% claims data.

list of regions by the DOJ (Appendix Table A.1); the remaining 4 are located in Texas and Louisiana within driving distance of strike-force HRRs.[13]

Third, as noted above, potentially fraudulent home health activity appears limited to one or two HRRs and are not characteristic of entire states. Harlingen, McAllen, and Dallas were all very high-spending HRRs in 2009 and 2010, but other Texas HRRs such as El Paso and Temple were much closer to the U.S. average. Similarly, Miami is an outlier even within Florida; the Fort Lauderdale HRR, adjacent to Miami, accounted for $1,175 in 2009, barely one-third the corresponding level in Miami, and 2009 home health expenditures in the Tallahassee HRR, $451, was below the U.S. average.

Fourth, there is a distinct rise and then decline in home health care expenditures, particularly for those targeted by the DOJ. The decline is likely to have been associated with two factors. The first is changes in policies enacted in response to potentially fraudulent activities. For example, until 2010, there were no restrictions on outlier payments, extra payments above the 60-day bundle allowed for unusually sick patients. While generally rare, the use of these outlier payments increased dramatically in Miami-Dade County, so much so that the county alone accounted for nearly half of U.S. home health outlier payments in 2009 (Benzio, 2010). In response, CMS restricted outlier payments for 2010 (Kim and Norton, 2015), leading to a particularly sharp reduction in home health care billing for the Miami HRR in 2010 as shown in Figure 1.

The other likely reason for the downturn is deterrence because of successful criminal and civil cases raising the perceived (and actual) probability of detection. For example, Leder-Luis (2020) found that Qui Tam or "Whistleblower" provisions for Medicare and Medicaid fraud led

[13] As noted in the notes to Table A.1, there is anecdotal evidence that the strike-force regions sometimes prosecuted cases located in these HRRs.

to a $6.80 specific deterrence effect per dollar of settlement; the general deterrence effect across the provider market is likely much larger. Despite the increased legal efforts to combat home health fraud, however, even in 2016 there remained considerable across regions with Miami and McAllen still exhibiting high home health care billing rates (Figure 1).

A key feature of the expansion in home health care expenditures was an increase in the *number* of home health care agencies in areas identified by the DOJ. In Figure 2, we show the number of home health agencies in the 9 HRRs as a ratio of the original number of agencies in 2002. In some cases, the number of agencies declines, as for example in New Orleans which experienced a decline in population after Hurricane Katrina in 2005, while Manhattan's number of agencies did not increase. But for 7 of the 9 regions the number of agencies grew rapidly, with a roughly 10-fold increase in Miami. This rapid increase may be the response to high incremental profits (particularly when services are not even provided) and low fixed costs (often just an office address and telephone), or it could be the result of a single business entity or family expanding the number of agencies. Unfortunately, tracing new agency formation is problematic as ownership is often disguised through shell corporations (Holly, 2020).

This section has established a diffuse pattern of home health care expansion that in equilibrium is consistent with a GSS model in which peer effects play an important role. We next turn to a more formal analysis of market structure identifying factors likely to contribute to a rapid diffusion in health care fraud.

III. A Model of Marginal Costs and Patient Referral Networks

Figures 1 and 2 raise questions about why expansion occurred so rapidly in some health care markets but not others. What is different about Miami and McAllen TX compared to Fort

Lauderdale or Temple TX? We hypothesize that in some regions, a bipartite *network structure* in place contributed to the diffusion of fraudulent billing activities spreading rapidly through the region.

Consider a model of home health care agency behavior in which the utility of the agency owner is a function of three objectives: Social benefits arising from patient treatment and care, the home health agency's profits, and the risks of conviction, fines, and possible imprisonment. The relative importance of each objective is assumed to vary across owners and markets. We follow Becker (1968) in writing expected utility of the home health care agency as a function of each:

(1) $$U_j = \alpha_j \phi(X_j) + \lambda_j P X_j - C(X_j) - \pi(X_j)\beta_j$$

In Equation (1), j denotes the individual home health agency,[14] $\phi(X_j)$ is the aggregate (dollar) social health benefits arising from the mix of k = 1, …, K inputs in agency j, X_j, while P is the vector of Medicare reimbursement rates for each of the k services. Financial cost $C(X_j)$ enters in the objective function, as does a more subtle form of cost; $\pi(X_j)$ is the probability of conviction depending on the legality of the collective activities summarized by X_j, while β_j is the penalty, which also can vary across agencies with regard to the perceived or actual penalty, defined broadly to include both economic and psychic adverse consequences.[15]

[14] Note that we focus on the home health agency rather than the physician, for two reasons. First, a reading of the legal cases suggests a central role for the agency owner(s), and while sometimes there was overlap between agency principals and the signing physician, often the physician was simply hired to refer and sign authorizations, or their authorization was forged. Second, preliminary analysis using physicians as "nodes" in network analysis suggested little evidence of systematic network effects.

[15] For simplicity we do not subscript π by region or time, but acknowledge that it may vary along these dimensions, for example if a DOJ strike force office is present; this regional variation is captured instead by variation in the penalty. As well, the penalty may differ depending on who is responsible for the fraud; physicians face a larger penalty than non-physician agency owners because they may lose their medical license.

The vector X_j describes the complex set of patient services (both type and quantity) provided by the home health agency. For example, suppose the kth element, X_{jk}, describes number of home health visits for well-documented patients (e.g., with strong medical justification); $X_{jk'}$ or the k′th element captures visits without documentation, while the k″th element measures the number of visits that are entirely falsified. The Medicare reimbursement rate for each visit would be the same but cost differences would arise (since undocumented visits can be expanded without limit, while the dollar marginal cost of a falsified visit is essentially zero). Most importantly the incremental probability of detection and penalty would be highest for the k″-type fraudulent visit.

We expect to see different input choices, outcomes, and Medicare reimbursement rates depending on the extent to which local home health agencies place more or less weight on patient health (through the parameter α_j), revenue and hence profits (λ_j), and on the perceived risks of criminal or civil penalties (β_j).[16] This can be seen from the first-order conditions:

(2) $$\alpha_j \frac{\partial \phi(X_j)}{\partial X_{jk}} + \lambda_j P_k - \frac{\partial C(X_j)}{\partial X_{jk}} - \frac{\partial \pi(X_j)}{\partial X_{jk}} \beta_j = 0$$

There are three polar cases that can be considered. The first is an agency that *maximizes social benefit*, so that $\alpha_j = 1$, $\lambda_j = 0$, and the risk of fraudulent detection is minimal so the third term is essentially zero. This type of agency corresponds to a (perhaps) not-for-profit firm with philanthropic goals. The second is an agency that *legally profit-maximizes*; it cares less about patient welfare but still wishes to avoid breaking the law; for them, $\alpha_j \approx 0$, $\lambda_j = 1$ but where they

[16] We ignore risk aversion and uncertainty here, although risk aversion would obviously make it less likely to engage in potentially fraudulent activity.

limit themselves to inputs such that for each k, the incremental risk of detection $\partial\pi / dX_{jk}$ is very small. The third polar case is one that *maximizes profits with potentially fraudulent strategies*; for this group $\alpha \approx 0$ *and* $\lambda = 1$ (so marginal revenue is no higher than marginal production cost) but where β_j or their perceived penalty (both monetary and psychic) is low.

In Figure 2, we show hypothetical supply curves for the three types of providers, where the price of a standardized home health care unit of service (as paid by Medicare) is shown as the horizontal black line, the horizontal axis is the number of patients, and the vertical axis is in dollar terms, either revenue, costs, or the dollar value of social benefits.[17] The green line is the (marginal) social benefit curve, where we assume that patients are lined up by their degree of appropriateness. The social optimum is given by Q_1 where the marginal social benefit is equal to the marginal cost of providing the service (the blue line, which corresponds to all entirely legal provisions of service). For the profit-maximizing but legally compliant agency, their output level is at Q_2, a point at which the marginal social benefit of the additional patient is well below the price the government pays, but where profits are maximized with minimal legal risks.

Finally, some agencies engage in fraudulent activities in which they access an entirely different set of inputs X including made-up services that cost nothing to provide, sharing and/or double-billing patients, and upcoding for expensive services that are not required; this in turn leads to a new and lower marginal cost curve, which we denote MC*. The lower marginal cost curve also brings with it increased risk of detection and penalty; the combined marginal cost of

[17] The price appears significantly above marginal cost; the U.S. General Accounting Office estimated margins of 16 to 18 percent in the early 2000s for home health agencies (Florida, 2007); currently Medpac (2020) estimates a marginal profit rate of 18 percent.

the incremental set of inputs X_j is denoted by the red dotted line, which intersects the price level at Q_3.

Note that knowledge about the potentially fraudulent activities captured by MC* is not universally shared; we hypothesize that it is discovered at the local level through informal networks reflected in (or generated by) patient-sharing by which agencies learn about strategies necessary to gain access to the frontier MC* technology.[18] While we measure patient sharing, we acknowledge that this isn't proof of information diffusion across agencies, but earlier evidence indicates patient sharing is closely associated with physicians seeking each other out for advice or one actively referring patients to the other (Barnett, 2011). Furthermore, we assume that the likelihood of agency A discovering a fraudulent strategy from agency B is a positive but diminishing function of the number of patients shared between the two agencies.[19] This means that the number of "edges" between home health care nodes, as would be captured by unipartite network measures, is inadequate to capturing the likelihood of dissemination across markets.

To see this, consider a hypothetical market with 10 home health agencies, of which initially just one is engaging in fraudulent activities. If this agency shared 5 patients with a second agency, the weighted sum of the edge-weights is 5, as it would be if it shared a single patient with 5 of the agencies. However, under the assumption of diminishing informational content associated with sharing multiple patients between two agencies, the latter case in which a single patient is shared across multiple agencies is predicted to lead to a greater market level of fraud compared to the former. This implies that a bipartite measure that captures the degree to

[18] If instead the MC* technology diffused by a more central mechanism – posting on the web, for example – we would not expect to find the sharp geographic variation observed in the data for the 2000s. An alternative explanation is that high-fraudulent regions just include more owners with preferences for potentially fraudulent activities, but this would not explain the change in activities and patient sharing since 2002.

[19] For a similar approach, see O'Malley et al. (2020).

which patients are shared with different home health agencies may capture a more rapid diffusion of fraudulent activity.

What are the aggregate implications of this rapid diffusion at the market level? A simple accounting exercise in the context of the model with three types of home health agencies would lead to the following measure of total home health expenditures, expressed in per-capita terms:

$$PQ_{jt} = P\left[\mu_{1jt}Q_{1jt} + \mu_{2jt}Q_{2jt} + \mu_{3jt}Q_{3jt}\right] \tag{3}$$

where μ_{lt} is the number of home health care agencies of type *l* relative to the Medicare population at time t.[20] In a dynamic sense, it is reasonable to assume that the relative fractions μ_{lt} will rise or fall depending on the relative profitability of the home health agency type. Given that legal (and even profit maximizing) home health agencies did not experience a fundamental change either in revenues, cost functions, or economies of scale, we assume that most increases in the number of home health agencies (unrelated to health needs) are the consequence of the diffusion in the knowledge and adoption of the MC* strategies coupled with high positive profits and low fixed costs (e.g., a phone number and office mailing address). We can write the change in the population share of potentially fraudulent activity from time t to t+1 for the (per capita) number of Type 3 agencies as:

$$\Delta\mu_3 = f(\varepsilon)(\Pi^* - \pi\beta_j) \tag{4}$$

where $(\Pi^* - \pi\beta_j)$ reflects the expected profitability, equal to the financial profits associated with fraudulent activity Π^* minus the risks of being caught and punished, $\pi\beta_j$, times the dynamic function $f(\varepsilon)$ that captures the speed of dissemination for information about the new fraudulent technology. We proxy ε with the BMIX index, although empirically we test for other network

[20] We consider a measure of home health care agencies per 10,000 Medicare enrollees below.

measures as well. We assume that changes in underlying health (as proxied by the age-sex-race-adjusted Medicare mortality rate) affects the quantity of services for Type 1 and 2 agencies, while ε is the primary determinant of the growth in Type 3 agencies.

IV. The BMIX Index

In this section, we develop a more formal measure of exposure to potentially fraudulent agencies and show how our bipartite approach differs from commonly used unipartite measures. Let B_{ih} be a binary variable that equals 1 if patient $h = 1, \dots, H$ received services from agency $i = 1, \dots, n$ within the year. In an undirected network, the bipartite mixture (BMIX) index for a region is defined as:

$$(5) \qquad BMIX = \frac{\sum_{i<k}^{n} A_{ik}}{\sum_{i<k}^{n} A_{ik} + \sum_{i=1}^{n} S_i}$$

where $A_{ik} = \sum_{h=1}^{H} B_{hi} B_{hk}$ is the number of instances when a patient receives support from both agencies *i* and *k* within a year and $S_i = \sum_{h=1}^{H} B_{hi} I(\sum_{i \neq k}^{n} B_{ik} = 0)$ is the total number of patients that receive support from agency *i* alone, where $I(event) = 1$ if event is true and 0 otherwise.

The index is a mix of a bipartite count of the number of single-source patients and a count of the number of times the edges in the projected network are affected owing to patients receiving services from multiple agencies. Unlike many unipartite measures, it is scale-free, thus making it useful in comparing markets of different sizes, and as we shall see, individual markets with rapid increase or decrease in the number of nodes (or agencies). In our application, the quadratic weighting of patients who receive services from multiple agencies and the use of information on the number of degree 1 (single source) patients will aid its predictive power.

Further insight into the BMIX index arises by defining a patient h attribute, denoted N_h, corresponding to the number of distinct agencies they received care from. Mathematically, $N_h = \sum_{i=1}^{n} B_{ih}$. Denote the number of patients with attribute value z by $d_z = \sum_{h=1}^{H} I(N_h = z)$.

Because $d_1 = \sum_{j}^{n} S_j$ and $\sum_{i<k}^{n} A_{ik} = \sum_{z=2}^{n} \binom{z}{2} d_z = \frac{1}{2}\sum_{z=2}^{n} z(z-1) d_z$ it follows that the BMIX index is also given by

$$(6) \qquad BMIX = \frac{\sum_{z=2}^{n} w_z d_z}{d_1 + \sum_{z=2}^{n} w_z d_z} = \frac{\sum_{z=2}^{n} w_z d_z}{\sum_{z=1}^{n} w_z d_z}$$

where $w_z = 1$ if $z = 1$ and $w_z = z(z-1)/2$ for $z > 1$, so that in general $w_z = I(z = 1) + z(z-1)/2$.

The expression in (6) shows that the BMIX index can be viewed as a network statistic of the bipartite network with patients and agencies as the two distinct sets of nodes. The numerator and denominator are weighted averages of the frequency distribution of the number of agencies patients received care from, a degree measure for the patient nodes in the bipartite patient-agency network. The weight for $z > 1$, $w_z = z(z-1)/2$, equals the number of edges in the patient-sharing network to which a patient who encounters z agencies contributes. Clearly, as the number of their agency encounters increases, the impact a patient has on the BMIX index increases quadratically. We are not aware of this measure having been previously developed in the network literature (although it is related to market overlap measures, as in Aryal et al., 2020); the presence of the indicator function in w_z for the $z = 1$ case makes the weights and function as a whole a mixture of patients that do ($z > 1$) and do not ($z = 1$) contribute to the network, thus making BMIX a mixture of two forms of information.

To illustrate the calculation of BMIX and the above points, suppose that two HRRs each have 10 agencies, with 20 patients in total. In the first agency, 9 beneficiaries receive services from exactly 2 of the agencies and 11 receive services from just one agency. In the second agency, 19 patients receive services from 1 agency and 1 patient receives services from all 10. The respective values of BMIX are:

$$BMIX_1 = \frac{2(2-1)/2 \times 9}{2(2-1)/2 \times 9 + 11} = \frac{9}{20} = 0.45$$

$$BMIX_2 = \frac{10(10-1)/2 \times 1}{10(10-1)/2 \times 1 + 19} = \frac{45}{64} = 0.70$$

The same number of services were provided to the same number of beneficiaries but the BMIX of the HRRs is very different because a single patient can more effectively serve as a "super-spreader" of potentially fraudulent strategies across all 10 agencies.

The BMIX index takes values ranging from 0 to 1. Therefore, without scaling, a regression coefficient for BMIX is interpreted as a change in the expected value of the outcome if BMIX equals 1 (all beneficiaries receiving care from two or more agencies so that $d_1 = 0$) compared to the counterfactual that it equals 0 (all beneficiaries receiving care from a single agency).

V. Peer Effects

The motivation for this paper stems from peer-effects, also known as social selection or contagion. For example, GSS has previously suggested that fraudulent financial strategies spread locally by learning from nearby criminals in a process consistent with peer-effects (also see

Zenou, 2003), and in Section II we hypothesized that a specific network structure is an element in which the diffusion of fraudulent billing activities can spread rapidly through a market. While we cannot establish causality, the presence of positive peer *associations* would suggest that the type of potentially fraudulent expenditures observed in home health expenditures may spread from agency-to-agency and thus be akin to peer-effects.

Our 15-year series of longitudinal data allows us to consider whether peer effects in the prior year may independently predict agency behavior in the current year. Because an agency (hereafter the "ego") may have multiple peer agencies, their combined influence on the ego can be quantified in a multitude of ways. In keeping with the premise for BMIX, we hypothesize that both the level of spending (outlying behavior) and the number of peers (reinforcement through multiple exposures) combine to impart influence. That is, being connected to more agencies will reinforce a willingness to spend higher, given the same average spending, and that exposure to multiple instances of high spending will have a greater impact than exposure to a single instance of extreme spending. We represent both an agencies number of peers (referred to as degree in network parlance) and the average spending of their peers as predictors in our peer association models.

Let Y_{ijt} denote the (single-source) expenditures of agency *i* in HRR j in year *t* and the adjacency matrix of a HRR network of HHAs by A_{jt} with *mn*th off-diagonal element $A_{jt,mn}$. We define a weight matrix, W_{jt} to be the row stochastic version of A_{jt} meaning that the rows sum to 1, implemented by dividing the elements on a row by their row sum; the row-sum for the ith row corresponds to the degree of agency *i* in HRR *j* and year *t*, denoted D_{ijt}. The diagonal elements of A_{jt} and W_{jt} are both equal to 0, so that the spending for shared patients of agency *i* is limited only to services received outside the ith agency. The product of W_{jt} and Y_{jt} yields a vector whose

ith element, $\bar{Y}_{ijt} = (WY)_{ijt}$, is the average spending of the agencies in HRR j with which agency i shares a network edge.

To adjust for factors unrelated to peer effects at the HRR and year level, we adjust for HRR-wide average spending of agencies that are isolated nodes in a given year (meaning that they have no edges with any other agencies), denoted $IsoSpend_{j(t-1)}$. The model already includes HRR fixed-effects and so this addition serves as a tighter control in that it would account for longitudinal HRR level factors that could be confounding the peer-effects (e.g., unmeasured time-varying HRR-level common causes). Because spending has a highly skewed distribution, we take the respective logs of ego agency spending, average peer agency spending, and HRR-wide average spending by isolate agencies. Only observations in which the focal agency was a non-isolate (and where there was at least one sole-source patient) were used in the estimation of the model.

For all home health agencies sharing at least one patient with another agency, the general model of interest is:

$$\begin{aligned}(7) \qquad \log(Y_{ijt}) &= \beta_{0t} + \beta_1 \log\left(Y_{ij(t-1)}\right) + \beta_2 \log\left(IsoSpend_{j(t-1)}\right) + \beta_3 \log\left(D_{ij(t-1)}\right) \\ &+ \beta_4 \log((WY)_{ij(t-1)}) + \beta_5 \log\left(D_{ij(t-1)}\right) \log((WY)_{ij(t-1)}) + \theta_j + \varepsilon_{ijt}\end{aligned}$$

where β_{0t} denotes year fixed-effects, β_1 is the association of the ego's prior year spending with its current spending, β_2 measures the association of isolate home health spending (e.g., for those not sharing patients) averaged across the HRR in the prior year with its current spending; this is designed to capture HRR-specific trends in patient health needs. The three key coefficients are β_3, the extent to which the number of peers of the agency (their network degree) is predictive of their spending in the following year, β_4, the extent to which the average spending across peer agencies is predictive of the agency's own spending in the following year,[21] and β_5, the modification of the peer average spending association by the focal agency's network degree. Finally, θ_j is a fixed-effect of the HRR to capture permanent differences in underlying health and other factors across HRRs and ε_{ijt} is a within-HRR error term.[22]

We consider four basic variants of Equation (7); one that excludes the own-agency lagged spending (e.g., setting $\beta_1 = 1$) and one that doesn't, one that assumes the interaction effect is null (e.g., setting $\beta_5 = 1$) and one that doesn't. To the extent that own-agency lagged spending already captures past peer associations, including the lagged effect is likely to bias downward the true peer association, but doing so may also partially mitigate homophily.[23]

VI. Data Preparation and Statistical Analyses

We build a beneficiary-sharing network for each HRR in each year. The nodes are home health agencies and the existence of an edge between two agencies indicates that at least one patient received care from both agencies during a calendar year. We also align the number of shared patients during the year with each such edge. Besides the bipartite BMIX index measure, we also construct three unipartite network measures by HRR and year chosen because they are classic measures with a theoretical and empirical basis for predicting fraudulent activity (Aven, 2015; Ferrara et al., 2014; O'Malley et al., 2020). The first is density, a commonly used measure capturing the fraction of potential connections among nodes (or agencies) with an edge. The

[21] We refrained from using models with peer-predictors from the current time-period as these would inflate the peer-effect due to agencies who share patients having also billed for those patients.
[22] Estimates using an HRR random-effects model are similar to the fixed-effects models presented below; thus we only present fixed effects coefficients.
[23] An alternative is to exclude $Y_{ij(t-1)}$ and account for repeated measurements across time through the error structure.

second is betweenness centralization quantifying heterogeneity in the extent that each agency intersects the information flow in the network.[24] Finally, we measured transitivity, a measure of network clustering quantifying "cliques" or unusually high density of edges among subsets of three nodes, as one might expect in fraudulent behavior.[25] While the BMIX index has a readily interpretable scale from zero to one, the others are more challenging and their distribution varies substantially with the number of nodes or agencies; for this reason we scale all four network measures by their standard deviations to facilitate interpretation. Because the measures other than BMIX vary mechanically with the number of agencies, we also include the number of agencies in 2002 in regressions involving these network measures.

Regression Models of Spending

The Medicare home health claims from the Dartmouth Atlas are used to create HRR-year level real per-enrollee expenditures in the fee-for-service population. These Atlas measures are based solely on residence, so if (e.g.) someone from the Rochester NY HRR winters in Tampa FL, their Tampa-based home health care spending would appear in the Rochester HRR measure. The primary outcome variable is the average price-adjusted, inflation-adjusted, and age-sex-race-adjusted payment per patient for an HRR and year. In regression models of per enrollee spending, we include the HRR/year mortality rates for the entire Medicare population on the right-hand side of the regression to adjust for differences in the health needs of the population.[26]

Our measures of home health care networks and other measures of utilization are slightly different; we count only home health visits by residents of an HRR that occur at a home health

[24] For example, a "ring" of agencies has a very low measure of betweenness centralization, while the hub node in a hub-and-spoke network has a very high one.
[25] We also considered the variance of several of these measures, but they also contained little explanatory power.
[26] Variability in home health spending is highly unlikely to have an impact on population-based mortality (e.g., McKnight, 2006).

care agency in that HRR. Each measure is constructed by year for each of the 306 HRRs from 2002 to 2016, leading to 4,590 total HRR/year observations. As a measure of fraudulent activity in the region, we also consider two other measures: (a) The regional growth in the *number* of home health agencies per 1,000 population relative to their 2002 frequency as a measure of profitability, and (b) whether the HRR was designated of interest to the Department of Justice strike forces, as noted above.

We use conventional cross-section time-series regression analysis, first using inflation-adjusted dollar values (per Medicare enrollee) but also considering per-enrollee spending in natural logs. We also consider whether our set of network measures in 2002 predict home health care utilization or number of home health agencies (per Medicare enrollee) in subsequent years 2003-16 conditional on *contemporaneous measures* of these same indices.[27]

Finally, and more directly, we ask whether network measures (or other measures) in 2002 could have predicted the subsequent growth in home health expenditures between 2002-09. Because network density and other measures are often related to the size of the network, we also include log of Medicare (fee-for-service) enrollees in the HRR and the number of "nodes" or agencies in the initial period 2002 as "intercepts" to adjust for different HRR market sizes. As a sensitivity test, we consider the reverse – did average home health spending in 2002 predict the change over time in the BMIX index from 2002-2009?

Peer Association Analysis

We also consider the analysis of individual home health agency expenditures within HRRs, as described in Section V. To guard against endogeneity from agencies that shared beneficiaries having their spending measure affected by the same patients, we restrict spending

[27] That is, in predicting (e.g.) 2012 home health spending in an HRR, we include both the contemporaneous 2012 BMIX measure, *and* the 2002 BMIX measure. Where appropriate, we cluster by HRR.

to the sample of beneficiaries that only receive care at a single agency ("single-source" beneficiaries). Therefore, the peer-agency regression coefficient is informed only by patients that did not contribute to the formation of the network; the estimated coefficient that obtains is likely to be a conservative measure of potentially fraudulent strategies. To aid the interpretation and comparability of parameter estimates across specifications, the resulting outcome and the peer-agency predictors were standardized to have a mean of 0 and a standard deviation of 1 across the subset of the dataset for which the original value of spending was positive. We used restricted maximum likelihood estimation, treating the ego agency as random effects to account for clustering and their HRRs as fixed effects to restrict the identification of estimates of peer associations to variation within markets. In a sensitivity analysis, the model was re-estimated with random effects for HRR; results were similar.

VII. Results

Table 1 presents summary statistics on home health care expenditures, for specific years 2002, 2009, and 2016, and network measures. We further considered these summary statistics for HRRs identified by the DOJ, and those not identified. For network measures, there is wide variability in the average value of the indices across regions, but the differences between the 21 HRRs identified by the DOJ and the other HRRs are modest for network density and transitivity. Betweenness centrality is about one-third lower, and BMIX about one-half higher, in the 21 DOJ HRRs compared to the remaining 285 HRRs. The largest differences between the DOJ and remaining HRRs are the number of home health agencies; there are 8 per 10,000 Medicare population in the DOJ-targeted HRRs, compared to 2.8 in the non-DOJ regions, and home health expenditures ($1152 compared to $479).

Recall that the BMIX index ranges between 0 and 1; as shown in Table 1 the mean is 0.15. Figure 4 shows graphically the distribution of BMIX by year using a box-and-whisker graph; it exhibits wide variability, while patient sharing rises across the distribution of HRRs during the peak expenditures of 2009-10. Miami is a consistent extreme outlier (labeled); Fort Lauderdale, Las Vegas, Houston, and Los Angeles were four other regions with high rates of BMIX.

The Structure of Patient-Sharing BMIX Networks

Figure 5 shows the Miami HRR in 2002 and 2009 as networks, while Figure 6 displays the patient-sharing network in Seattle (a low-growth region) during the same years. The 2009 plot presents just the most connected nodes with the number of displayed agencies equaling the total number of agencies in the 2002 network. The nodes are colored with red (most), blue and green (least) corresponding to the number of beneficiaries who only receive care from that agency. Miami illustrates a fundamental change in the degree of patient sharing – a shift from little patient sharing in 2002 (green nodes) to common sharing in 2009 (red nodes) during this period, but Seattle remains relatively stable.

The association between the BMIX index and spending measures can be seen by sorting HRRs into deciles by their BMIX measure, either in 2002 or in 2009. In Figure 7, Panel A shows a modest positive association between the 2002 BMIX index and 2002 home health expenditures per enrollee; the correlation becomes much stronger in 2009 (Panel B) particularly for the very top BMIX decile. The 2002 BMIX index predicts the subsequent growth in home health spending between 2002-09 (Panel C) and the corresponding growth in the number of home health agencies per 10,000 enrollees (Panel D). While most of the agency growth and expenditure growth is associated with the regions corresponding to the top decile of the BMIX index, there appears to be a broader association between BMIX and home health expenditures across all deciles, particularly in 2009.

Spending Regression Results

The regression model in Table 2A allows for other network measures as explanatory variables in predicting home health care expenditures. As noted above, each of the four network measures (starred) has a standard deviation of 1.0 and a mean of zero; the interpretation of each of these coefficients is the change in the dependent variable with respect to a one-standard-deviation change in the independent variable.

The first two columns are least squares regressions both unweighted and weighted by the number of Medicare enrollees; all regressions include controls for year and are clustered by HRR. In the first column, a one-standard-deviation increase in BMIX is predicted to increase home health care spending by $173, or 33 percent of average home health spending; other network measure coefficients are smaller in magnitude or negative.[28] A higher mortality rate is associated with higher home health spending; a one-standard deviation increase is predicted to increase home health spending by $118;[29] weighted regression coefficients are similar. Column 3, for years 2003-16, includes both contemporaneous and 2002 levels of (standardized) network measures. Once again, BMIX enters significantly, with a slightly larger coefficient for the 2002 value relative to the contemporaneous BMIX measure; their combined impact, corresponding to

---

[28] While the theoretical sign of the BMIX coefficient is positive, one can construct theories for why the other network measures would exhibit a negative rather than positive association. For example, one might argue that higher triadic closure (higher transitivity) is an indication of greater coordination of unnecessary beneficiary sharing among "like" agencies (perhaps those owned centrally). But higher transitivity could be explained as well by coordination or geographic proximity within a region; it would be unusual if agencies A and B shared patients and agencies A and C shared patients, but B and C did not. The question is then whether such sharing patterns should predict higher home health care expenditures. Likewise, one might argue that a network with greater centralization is akin to a hub and spoke network whereby one agency dominates patient sharing such that it "polices" the others and thus guards against fraudulent activity.

[29] The coefficient, $203.7, times the standard deviation, from Table 1 (.579).

a permanent increase in BMIX during the entire period (196.5 = 86.4+110.1) is larger than the coefficient in Column 1. In the HRR-fixed-effects models, the association between BMIX and spending is much diminished, suggesting that spending does not track year-to-year with BMIX (or any of the other variables); we consider longer-term growth in home health care spending below in Table 2D.

Table 2B displays similar estimated results for the log of home health care spending by HRR and by year, with a one-standard deviation rise in BMIX predicted to increase log home health spending by nearly one third (Columns 1 and 2), but (as in Table 2A) with much smaller predicted effects in the HRR fixed effects models. Table 2C considers similar regressions, but with the number of home health care agencies (per 10,000 Medicare population) – a measure of profitability of home health care in a given region -- as the dependent variable. Even when controlling for the number of agencies in 2002, the regressions suggest a significant positive impact of BMIX on the expansion of agencies. In this Table, the network density, betweenness centrality, and network transitivity measures are more closely associated with the number of agencies, although in part it may be because conventional network measures are often sensitive to the number of nodes in the network structure.

Table 2D provides estimates of changes in home health expenditure and number of agencies between 2002-2009; except for the corresponding change in mortality during the same period (to adjust for changes over time in underlying health), we used only information known in 2002. Column 1 shows that the BMIX in 2002 is strongly predictive of growth, with a one-standard-deviation difference predicting $180 more rapid growth (or 68 percent of the average increase between 2002 and 2009), but with smaller effects ($85, or 32 percent of average growth) when lagged home health spending from 2002 is included in the regression (Column 2).

With logged home health spending changes (Column 3), the coefficient on BMIX is .088 (or 20 percent of the log change) with the dependent variables included. The reverse association does not hold; a regression of the change in the BMIX between 2009 and 2002 on the initial level of home health spending in 2002 shows a *negative* (and barely significant) correlation.

Finally, Table 2D shows changes between 2002 and 2009 in two additional outcomes. Column 4 shows a strong positive association between the 2002 BMIX and the change in the number of home health care agencies, while a linear prediction model suggests that the likelihood that a given HRR would be designated one of the 21 strike-force HRRs increases by 8.6 percent when the BMIX is one standard deviation higher; both results are highly significant.

Peer Association Analysis Results

The results of the peer-associated analyses are displayed in Table 3, based on 1.54 million shared home health care patients across home health agencies in the 306 HRRs. We use the Bayesian Information Criterion (BIC), also known as the Schwarz criterion (Schwarz 1978), and marginal R^2 (the random effects are considered part of the error-term as opposed to part of the model fit) to compare the fitted models.[30] We find the specifications with the interaction between average peer spending and number of peer agencies (i.e., degree) included yields a superior fit than when this is excluded; the marginal R^2 also demonstrates an advantage of including the interaction term but on a much more restrictive scale.[31] Therefore, we focus

[30] Because the models have the same number of predictors and are estimated on the same data set, all likelihood-based model comparison procedures are essentially equivalent.

[31] Because we estimate a linear model, the interpretation of the effects as unit-changes in log-predictors in relation to standardized log-spending in the following year is straight-forward and is not subject to the considerations for nonlinear models in Ai and Norton (2003). In the case of models (2) and (4), the interpretation of a unit change in lagged log average peer spending on the focal agencies spending is more complicated and depends on the values of other variables in the model. This is seen by the fact that the partial derivative of Y_{ijt} with respect to $(WY)_{ij(t-1)}$ in the general model in (7) depends on the exponential of the right-hand-side (RHS) multiplied by $(\beta_4 + \beta_5 \log(D_{ij(t-1)}))/(1 + (WY)_{ij(t-1)})$, clearly a nonlinear expression involving all RHS variables and parameters.

primarily on interpreting the results of the analyses of the models that allow the association of peer-agency spending to be modified by the degree of the focal agency.[32] We normalize both variables involved in the interaction so that their standard deviation is 1.0.

In the absence of the interaction, the dominant network-related predictor is lagged log degree; in the model without lagged ego spending (e.g., the lagged dependent variable) the estimated coefficient is 0.327 (standard error 0.002) and in the model with lagged ego spending the estimated coefficient is 0.029 (standard error 0.003). In both of these models, lagged log peer average spending is statistically non-significant. We believe that these two pairs of estimates likely bracket the true peer association; the former is likely biased upward because of homophily or unmeasured common causes acting contemporaneously across an HRR, while the latter biased downward because the spending measure is limited to "isolate" (unshared) patients, and does not capture the dynamics by which peer associations in past years are already reflected in year $t - 1$ ego spending measures.

With the addition of the interaction, the overall impact of lagged log peer average spending amplifies; in the model without lagged ego spending the main and interaction estimated coefficients are 0.036 (0.003) and 0.036 (0.002), respectively, and in the model with lagged ego spending they are 0.021 (0.004) and 0.009 (0.002).[33] Considering the interactive terms in Column 4, the model predicts that when average logged spending is at the 90th percentile (1.28 standard-deviations above the mean), the association between a one-standard deviation increase in the logged number of peers and subsequent log spending by the ego is 0.033; the

[32] We also focus on the models with random-effects for HHA and fixed-effects for HRR, noting the trivial differences in estimates between the model when HHR is also specified as a random-effect.
[33] The finding that a combination of a network summary measure about the focal agency (degree) and the network-averaged outcome of their peer agencies combine as predictors in the best-fitting model emulates the general structure and results found in O'Malley et al (2020).

corresponding estimate for a one-standard-deviation increase in logged average spending at the 90th percentile for logged number of peers is 0.030. Finally, the prediction associated with a simultaneous increase in logged average spending and logged average number of peers from their means to their 90th percentiles is 0.065.

In sum, we have established that even *within* HRRs, home health agencies sharing patients with a greater number of other agencies or with high-spending other agencies were more likely to increase patient expenditures in the following year. We also explored extending the model in Equation (7) to allow the peer-associations to be modified by the lagged BMIX of the HRR. While not reported, we found that in HRRs with a higher BMIX, the peer association coefficients were smaller in magnitude, suggesting diminishing returns to additional information about agencies with which the ego agency shares patients.

VIII. Discussion and Conclusion

It is well established that there are wide geographic variations in the diffusion of highly effective health care (Coleman et al., 1966; Jencks et al., 2003; Skinner and Staiger, 2007) and newly developed cancer drugs (Agha and Molitor, 2018); much less is known about the diffusion of ineffective or potentially harmful use of potentially fraudulent health care. In this paper, we have studied a rapid increase in billing for Medicare home health care expenditures in some but not all regions of the U.S. during the 2000s. These billing increases cannot be explained by changing health needs, nor can they be explained by the substitution of inpatient for home health care.[34] Instead they appear largely the consequence of widespread fraudulent behavior which in

[34] Using price-adjusted spending for 2003-09 at the HRR level, the correlation coefficient between the change in home health care spending and the change in inpatient and nursing home case was 0.237 (p < .01). McKnight (2006) showed that when home health reimbursements were capped, there was no increase in hospital use.

turn attracted specific Department of Justice strike-force offices located in areas with rapid increases in fraudulent behavior.

Using a theoretical model as guidance of fraudulent billing in which the potential gains from such activity outweigh penalties of legal convictions, we develop a novel bipartite mixture network index, the BMIX, that predicts the diffusion across home health agencies of fraudulent billing as reflected in the placement of DOJ strike force offices. Commonly-used unipartite network measures such as density, betweenness-centrality, and transitivity were much less predictive of this rapid increase in home health spending. But the 2002 BMIX, measured at the outset of the sharp rise in home health care spending, was predictive of the growth in *subsequent* home health care expenditures.[35] The reverse did not hold, however, higher 2002 spending predicted slightly lower BMIX growth.

The BMIX index is related to, but distinct from, the idea of fragmentation, or the receipt of care from multiple physicians. Fragmentation involves tradeoffs; more physicians can in theory provide greater specialization, albeit with rising coordination costs (as in Becker and Murphy, 1992). However, the empirical evidence suggests that fragmentation raises costs and reduces quality (Agha et al., 2019). Yet home health agency patient-sharing differs from fragmentation or physician patient-sharing in that the agency is supposed to be coordinating the patient's care, thus the value of involving a second home agency is unclear.[36]

The model suggests why in some regions, home health care fraud might diffuse more rapidly than in other regions, but we still cannot say why Miami and McAllen Texas (and not Fort Lauderdale or Temple Texas) were the regions to experience such rapid change. One potential explanation for why Miami's home health care sector grew so rapidly was relaxed

[35] We recognize that other bipartite measures (e.g., Opsahl, 2013) may also predict diffusion.
[36] The analogy might be when a patient is enrolled in multiple managed care organizations.

state-level regulation in Florida, yet some regions in Florida experienced only average growth in home health care. Another possibility is the idea of hysteresis or long-standing experience in fraudulent activities. For example, in the mid- and late-1990s, there was a similar and short-lived outbreak of potentially fraudulent home health services in the rural South and West (Vandenburgh, 2005). Dartmouth Atlas data show that in 1995 McAllen Texas was the highest-spending region for home health services, with Miami ranked barely in the top decile of HRRs. Yet Miami during this period led the country with regard to expenditures for durable medical equipment; by 2001 it was spending 6 times the U.S. average for potentially fraudulent expenditures on (e.g.) motorized wheelchairs.[37] Thus Miami entrepreneurs could have simply switched from durable medical equipment to home health care during the 2000s. Hysteresis, however, cannot explain why Chicago (ranked 133rd in 1995) should have exhibited such unusually rapid increases in home health expenditures during the 2000s.

Collectively and individually, the pattern of spending for home health agencies *within* HRRs support the presence of peer-associations. The coefficient estimates are consistent with provider communication that, like Barnett et al. (2012) occurs through sharing of patients across agencies (also see O'Malley et al. 2014; 2020).[38] These results are also consistent with a model in which rapid market growth occurs when knowledge of a new business model with very low marginal costs – such as fictitious services for patients – and low fixed costs spreads across a market, as is suggested by the 10-fold increase in the number of home health agencies in Miami over just a few years.

[37] www.dartmouthatlas.org, accessed February 18, 2021. Durable medical equipment fraud in South Florida played a central role in Carl Hiaasen's novel "Bad Monkey" (2013).
[38] Goel (2020) suggests that these spillover effects may extent even beyond state borders, a result consistent with our evidence from HRRs which often cross state lines.

We acknowledge two limitations of the analysis. The first is that we cannot measure fraud directly because agencies are understandably reticent about their potentially illegal behavior, and because those successfully charged are only the tip of the iceberg; most fraudulent (or gray-area) activities are difficult to detect. While we infer fraudulent behavior using a variety of approaches including the number of home health agencies and the presence of a DOJ strike-force in the HRR, we recognize that some of the three-fold increase in home health spending in Miami between 2002 and 2009 could have been legitimate. Second, while we established that the BMIX index is theoretically consistent with a model of diffusion, and is predictive of subsequent growth in home health expenditures, we cannot prove causality. The BMIX index is not likely to be capturing unmeasured health effects – it is uncorrelated with mortality[39] – but patient sharing patterns could be symptomatic of past or current fraudulent activity which in turn lays the groundwork for future fraudulent behavior. Still, at a minimum the BMIX index is predictive of future fraudulent behavior, suggesting value in machine-learning approaches to unearthing Medicare fraud (Bauder et al., 2017).

There are a variety of other applications for bipartite network measures that can potentially capture models of diffusion and infection, for example patterns of staff-sharing across nursing homes leading to rapid diffusion of COVID-19 infections among nursing home patients (Chen et al., 2020). One requires relatively recent data to calculate network structures useful in predicting future behavior, but government agencies should be able to acquire the claims data needed to compute networks of home health care or of other relevant organizations with only a

[39] A regression of BMIX on mortality with year fixed effects yields a coefficient of .003 (standard error, .007); similar results are obtained with year and HRR fixed effects. This means that any correlation between BMIX and unmeasured health is independent of mortality.

few months lag.[40] Furthermore, the government may have access to records of past occurrences of fraud that could be used to train a predictive machine-learning model to make optimal BMIX-based predictions.

Network analysis may also be used to test whether fraud more generally is "contagious" beyond financial settings (Dimmock et al., 2018). For example, Howard and Desai (2020) document hospitals (or hospital systems) accused of providing unnecessary stents (percutaneous coronary interventions) for their patients. Of the 16 systems in their study, 10 were located in just 4 states: Pennsylvania, Maryland, Ohio, and Kentucky, a finding consistent with networks of interventional cardiologists within and across hospital systems. In other cases, health care fraud may diffuse through consultants or device manufacturers advising hospitals to follow overly aggressive (and ultimately illegal) billing practices (Silverman and Skinner, 2004; U.S. Department of Justice, 2014). In these cases, networks might not exhibit the sharp geographic clustering observed in home health care fraud, but a network properly constructed could still exhibit the bipartite patterns we see in home health care. While future research is needed to test whether measures such as the BMIX can predict diffusion in other settings, we believe there is a strong basis for the use of network analysis in the analysis of health care fraud and market dynamics.

[40] Lee and Skinner (2021) consider several other potential policy changes that would make it easier for the federal government to pursue home health care fraud.

References

Ai, Chunrong, and Edward C. Norton. "Interaction terms in logit and probit models." *Economics Letters* 80, no. 1 (2003): 123-129.

Agha, Leila, and David Molitor. "The local influence of pioneer investigators on technology adoption: evidence from new cancer drugs." *Review of Economics and Statistics* 100, no. 1 (2018): 29-44.

Agha, Leila, Brigham Frandsen, and James B. Rebitzer. "Fragmented division of labor and healthcare costs: Evidence from moves across regions." *Journal of Public Economics* 169 (2019): 144-159.

Aryal, Gaurab, Dennis J. Campbell, Federico Ciliberto, and Ekaterina Khmelnitskaya. "Common-subcontracting and multimarket contact in the airline industry." University of Virginia Working Paper (2020).

Aven, Brandy L. "The paradox of corrupt networks: An analysis of organizational crime at Enron." *Organization Science* 26, no. 4 (2015): 980-996.

Baker, Wayne E., and Robert R. Faulkner. "Diffusion of fraud: Intermediate economic crime and investor dynamics." *Criminology* 41, no. 4 (2003): 1173-1206.

Barnett, Michael L., Bruce E. Landon, A. James O'Malley, Nancy L. Keating, and Nicholas A. Christakis. "Mapping physician networks with self-reported and administrative data." *Health Services Research* 46, no. 5 (2011): 1592-1609.

Barros, Pedro, and Gisele Braun. "Upcoding in a National Health Service: the evidence from Portugal." *Health Economics* 26, no. 5 (2017): 600-618.

Bauder, Richard, Taghi M. Khoshgoftaar, and Naeem Seliya. "A survey on the state of healthcare upcoding fraud analysis and detection." *Health Services and Outcomes Research Methodology* 17, no. 1 (2017): 31-55.

Becker, Gary S. "Crime and punishment: An economic approach." In Gary S. Becker and William M. Landes (eds.) *The Economic Dimensions of Crime*, pp. 13-68. Palgrave Macmillan, London (1968).

Becker, Gary S., and Kevin M. Murphy. "The division of labor, coordination costs, and knowledge." *The Quarterly Journal of Economics* 107, no. 4 (1992): 1137-1160.

Benzio, Brooke. “Fee-for-Disservice: Medicare fraud in the home healthcare industry,” *Annals of Health Law* Volume 19 Issue 1 (Article 41) (2010).

Chen, M. Keith, Judith A. Chevalier, and Elisa F. Long. “Nursing home staff networks and COVID-19.” No. 27608. National Bureau of Economic Research (2020).

Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. "The association between income and life expectancy in the United States, 2001-2014." *JAMA* 315, no. 16 (2016): 1750-1766.

CMS, Center for Medicare and Medicaid Services, Medicare and Home Health Care (2003). https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/Downloads/HHQIHHBenefits.pdf (Accessed February 21, 2021).

Coleman, James, Elihu Katz, and Herbert Menzel. *Medical Innovation: A Diffusion Study* Indianapolis, The Bobbs-Merrill Company, Inc. (1966).

Cutler, David, Jonathan S. Skinner, Ariel Dora Stern, and David Wennberg. "Physician beliefs and patient preferences: a new look at regional variation in health care spending." *American Economic Journal: Economic Policy* 11, no. 1 (2019): 192-221.

Dafny, Leemore S. "How do hospitals respond to price changes?" *American Economic Review* 95, no. 5 (2005): 1525-1547.

Dimmock, Stephen G., William C. Gerken, and Nathaniel P. Graham. "Is fraud contagious? Coworker influence on misconduct by financial advisors." *The Journal of Finance* 73, no. 3 (2018): 1417-1450.

Ferrara, Emilio, Pasquale De Meo, Salvatore Catanese, and Giacomo Fiumara. "Detecting criminal organizations in mobile phone networks." *Expert Systems with Applications* 41, no. 13 (2014): 5733-5750.

Florida, State of. The Florida Senate Interim Project Report 2008-135 November (2007). http://archive.flsenate.gov/data/Publications/2008/Senate/reports/interim_reports/pdf/2008-135hr.pdf

GAO, Medicare and Medicaid: CMS needs to fully align its antifraud efforts with the fraud risk framework, GAO-18-88 (Washington, D.C.: Dec. 5, 2017).

Glaeser, Edward L., Bruce Sacerdote, and Jose A. Scheinkman. "Crime and social interactions." *The Quarterly Journal of Economics* 111, no. 2 (1996): 507-548.

Goel, Rajeev K. "Are health care scams infectious? Empirical evidence on contagion in health care fraud." *Managerial and Decision Economics* (2020).

Gottlieb, Daniel J., Weiping Zhou, Yunjie Song, Kathryn Andrews, Jonathan Skinner, and Jason M. Sutherland. "Prices don’t drive regional Medicare spending variations." *Health Affairs* 29, no. 3 (2010): 537-543.

HHS/GAO. The Department of Health and Human Services and The Department of Justice Health Care Fraud and Abuse Control Program Annual Report for Fiscal Year 2016 (2017). https://oig.hhs.gov/publications/docs/hcfac/FY2016-hcfac.pdf

Holly, Robert. “Home health owners pay price for $38 million ‘Empire of Fraud’.” Home Health Care News, (January 1, 2020). https://homehealthcarenews.com/2020/01/home-health-owners-pay-price-for-38-million-empire-of-fraud/

Howard, David H., and Nihar R. Desai. "US False Claims Act investigations of unnecessary percutaneous coronary interventions." *JAMA Internal Medicine* 180, no. 11 (2020): 1534-1536.

Hyman, David A. "Health care fraud and abuse: market change, social norms, and the trust “reposed in the workmen”." *The Journal of Legal Studies* 30, no. S2 (2001): 531-567.

Imperato, Gabriel, An overview of pervasive allegations in hospice and home health fraud cases, Broad and Cassel, Fort Lauderdale, FL. (2017) https://www.hcca-info.org/Portals/0/PDFs/Resources/Conference_Handouts/Regional_Conference/2017/orlando/Imperatohandout2.docx

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* 50, no. 5 (2015): 1606-1627.

Jencks, Stephen F., Edwin D. Huff, and Timothy Cuerdon. "Change in the quality of care delivered to Medicare beneficiaries, 1998-1999 to 2000-2001." *JAMA* 289, no. 3 (2003): 305-312.

Leder-Luis, Jetson. “Whistleblowers, private enforcement, and Medicare fraud,” Working Paper, Boston University (2020).

Lee, Stephen, and Jonathan Skinner. “Reforming Health Care Coverage to Reduce Fraud,” (2021). https://onepercentsteps.com/policy-briefs/reforming-home-health-care-coverage-to-reduce-fraud/

McKnight, Robin. "Home care reimbursement, long-term care utilization, and health outcomes." *Journal of Public Economics* 90, no. 1-2 (2006): 293-323.

Medpac (Medicare Payment Advisory Commission). Report to the Congress: Medicare Payment Policy. (March 2020). http://www.medpac.gov/docs/default-source/reports/mar20_medpac_ch9_sec.pdf

Nash, Rebecca, Martin Bouchard, and Aili Malm. "Investing in people: The role of social networks in the diffusion of a large-scale fraud." Social Networks 35 No. 4 (2013): 686-698.

O'Malley, A. James, Erika L. Moen, Julie PW Bynum, Andrea M. Austin, and Jonathan S. Skinner. "Modeling peer effect modification by network strength: The diffusion of implantable cardioverter defibrillators in the US hospital network." *Statistics in Medicine* 39, no. 8 (2020): 1125-1144.

O'Malley, A. James, Felix Elwert, J. Niels Rosenquist, Alan M. Zaslavsky, and Nicholas A. Christakis. "Estimating peer effects in longitudinal dyadic data using instrumental variables." *Biometrics* 70, no. 3 (2014): 506-515.

Opsahl, Tore. "Triadic closure in two-mode networks: Redefining the global and local clustering coefficients." *Social Networks* 35, no. 2 (2013): 159-167.

Sah, Raaj K. "Social osmosis and patterns of crime." *Journal of Political Economy* 99, no. 6 (1991): 1272-1295.

Schwarz, Gideon E. "Estimating the dimension of a model", *Annals of Statistics*, 6, no. 2 (1978): 461–464.

Silverman, Elaine, and Jonathan Skinner. "Medicare upcoding and hospital ownership." *Journal of Health Economics* 23, no. 2 (2004): 369-389.

Skinner, Jonathan. "Causes and consequences of regional variations in health care." In *Handbook of Health Economics*, vol. 2, Elsevier (2011): 45-93.

Skinner, Jonathan, and Douglas Staiger. "Technology diffusion and productivity growth in health care." *Review of Economics and Statistics* 97, no. 5 (2015): 951-964.

U.S. Department of Justice. "Washington-Based Medical Device Manufacturer to Pay up to $5.25 Million to Settle Allegations of Causing False Billing of Federal Health Care Programs" (2014). https://www.justice.gov/opa/pr/washington-based-medical-device-manufacturer-pay-525-million-settle-allegations-causing-false

Vandenburgh, Henry. "Home health care fraud in the 1990s: Did region play a role?" *Journal of Economic Crime Management* 13, no. 1 (2005): 1-13.

Welch, H. Gilbert, David E. Wennberg, and W. Pete Welch. "The use of Medicare home health care services." *New England Journal of Medicine* 335, no. 5 (1996): 324-329.

Weng, Weifeng, Jessica Van Parys, Rebecca S. Lipner, Jonathan S. Skinner, and Brenda E. Sirovich. "Association of Regional Practice Environment Intensity and the Ability of Internists to Practice High-Value Care After Residency." *JAMA Open* 3, no. 4 (2020): e202494-e202494.

Woolf, Steven H., and Heidi Schoomaker. "Life expectancy and mortality rates in the United States, 1959-2017." *JAMA* 322, no. 20 (2019): 1996-2016.

Zenou, Yves. "The spatial aspects of crime." *Journal of the European Economic Association* 1, no. 2-3 (2003): 459-467.

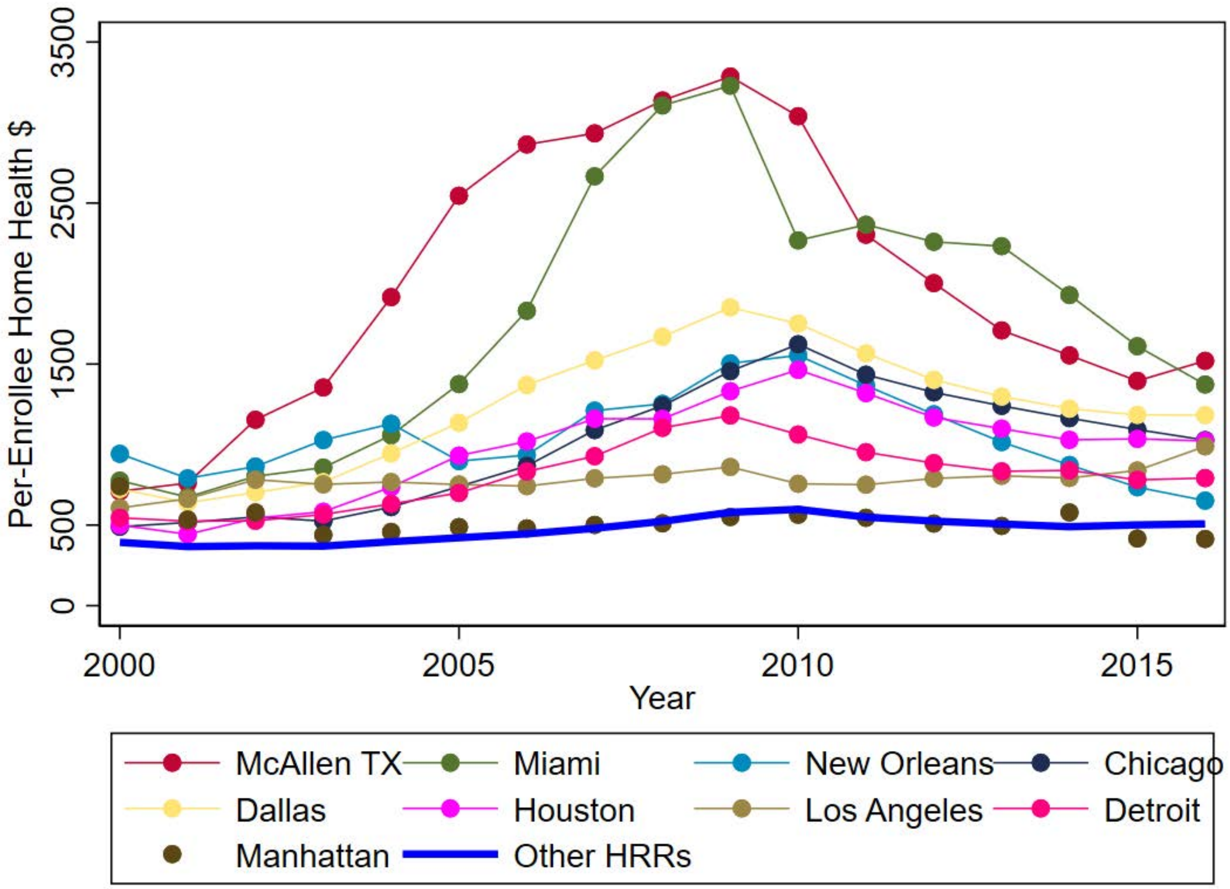


Figure 1: Per-Enrollee Home Health Expenditures, 2000-2016, by Selected Region
Note: Selected regions from the 21 HRRs selected by the Department of Justice as a location for (or area of interest of) their fraud strike forces. "Other" is the weighted average of the other 285 HRRs. All Expenditures in 2016 dollars.

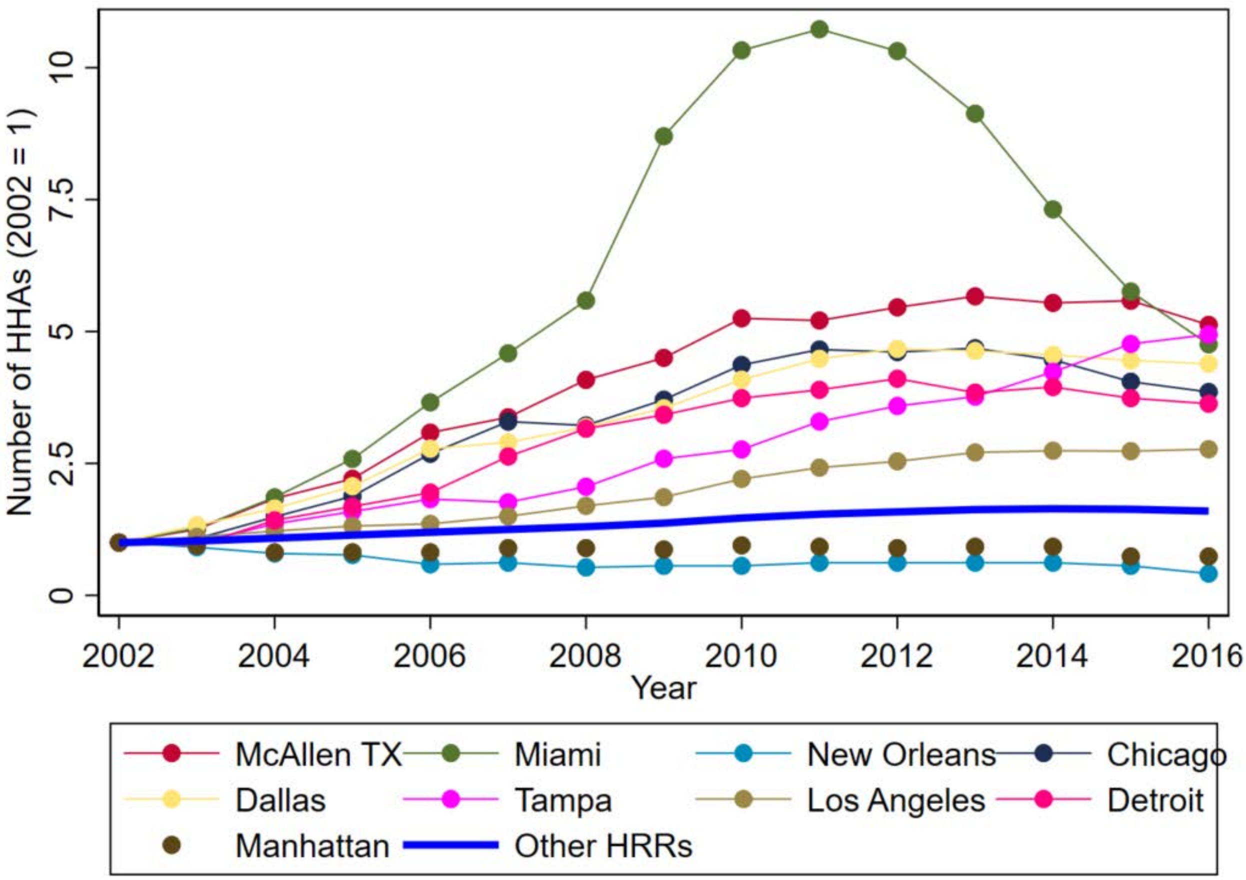


Figure 2: The Number of Distinct Home Health Care Agencies Relative to 2002, by Selected Region

Note: Selected regions from the 21 HRRs selected by the Department of Justice as a location for (or area of interest of) their fraud strike forces. "Other" is the weighted average of the other 285 HRRs. All measures are relative to the initial number of HRRs in 2002; thus New Orleans likely experienced a decline in the number of home health agencies because of a decline in population after Hurricane Katrina in 2005.

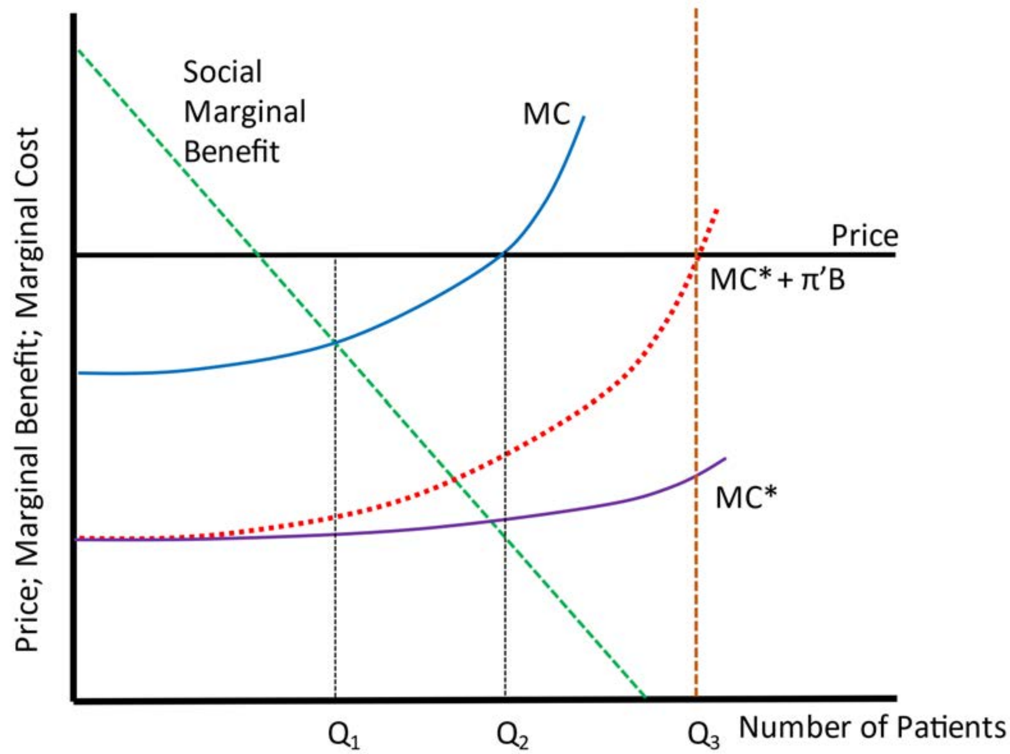


Figure 3: Marginal Cost/Supply Curves for 3 Types of Home Health Care Agencies
This figure characterizes the different output decisions for three different types of home care agencies. The first type is one that maximizes social welfare, setting marginal social benefits (the downward sloping green curve) equal to marginal cost at Q_1. The second type of agency maximized profits (legally) by setting marginal cost equal to price, at quantity Q_2. Finally, the third type of agency faces a very low marginal cost curve (MC*) but an ex ante risk of prosecution (the difference between the red dotted line and the MC* line); this third agency maximizes their objective function at Q_3.

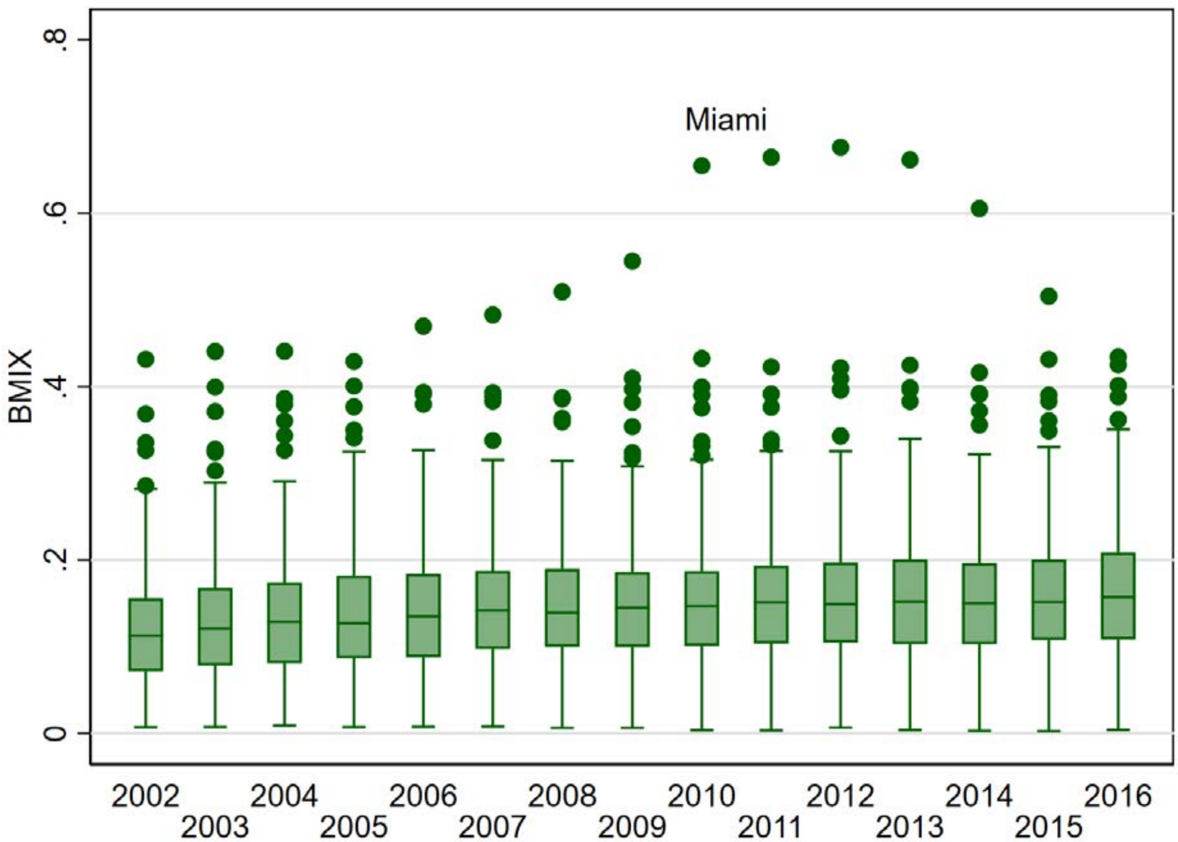
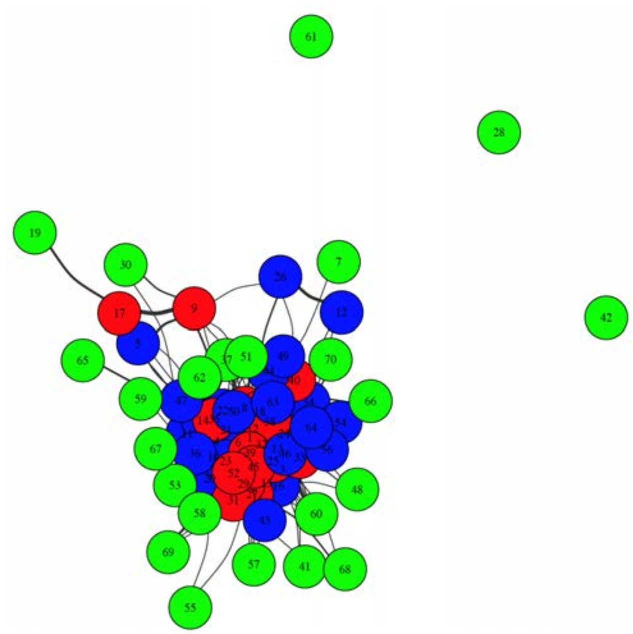


Figure 4: Box and Whisker Plots of the BMIX Index by HRR and by Year, 2002-16
The shaded bar represents the interquartile range (25th to 75th percentile) with the median marked by the horizontal bar. The "whiskers" are the 95th percentile with individual dots as outliers. Miami appears as an outlier in multiple years (as labeled). In 2010, the HRRs with the 5 highest BMIX measures were Miami (.65), Las Vegas (.43), Fort Lauderdale (.40), Houston (.39), and Los Angeles (.37).

Panel A: 2002



Panel B: 2009

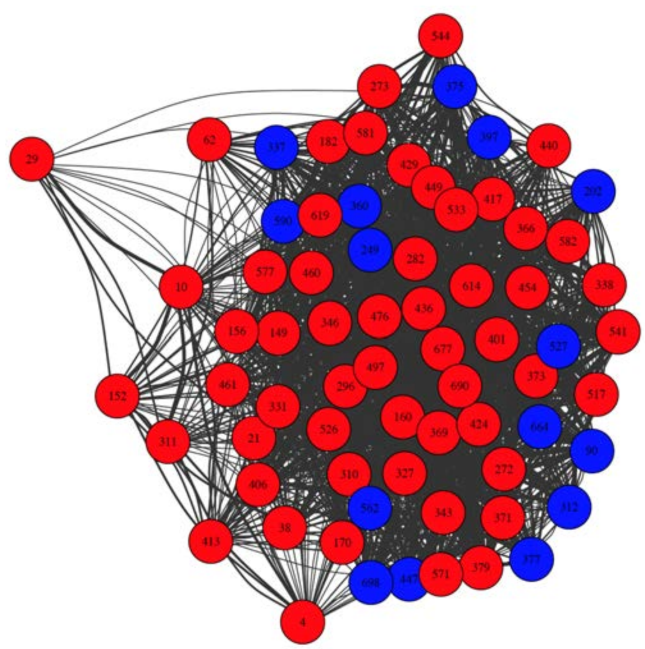
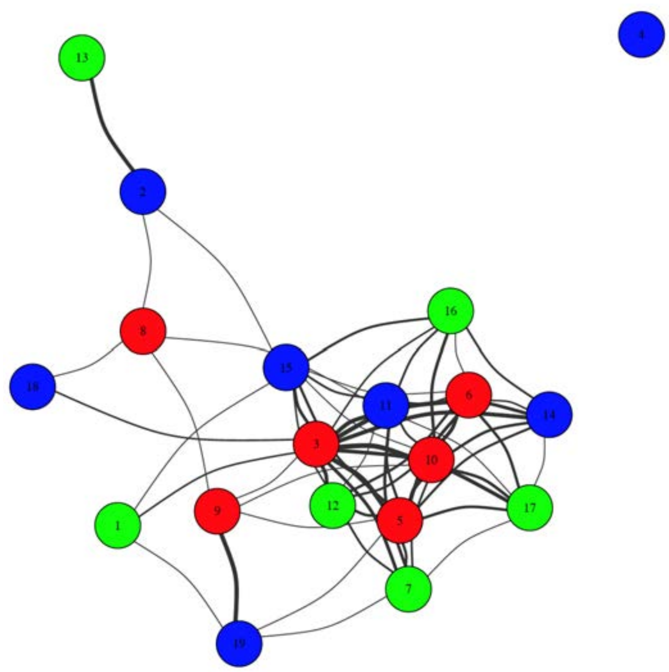


Figure 5: Network plots for the Miami HRR (Panel A: 2002, Panel B: 2009).
The 2009 plot is restricted to the most connected agencies of number equal to the total number of agencies in 2002. The nodes are colored with red (most), blue and green (least) corresponding to the number of beneficiaries who only receive care from that agency.

Panel A: 2002



Panel B: 2009

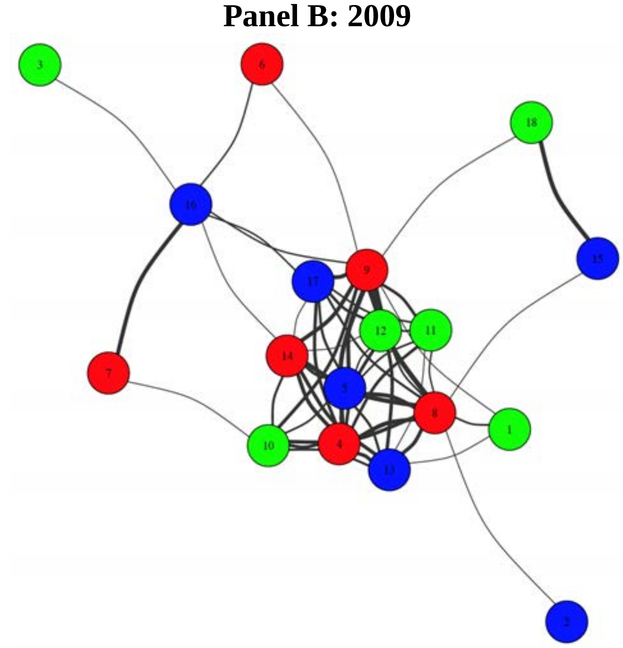


Figure 6: Network plots for the Seattle HRR (Panel A: 2002, Panel B: 2009)
The nodes are colored with red (most), blue and green (least) corresponding to the number of beneficiaries who only receive care from that agency.

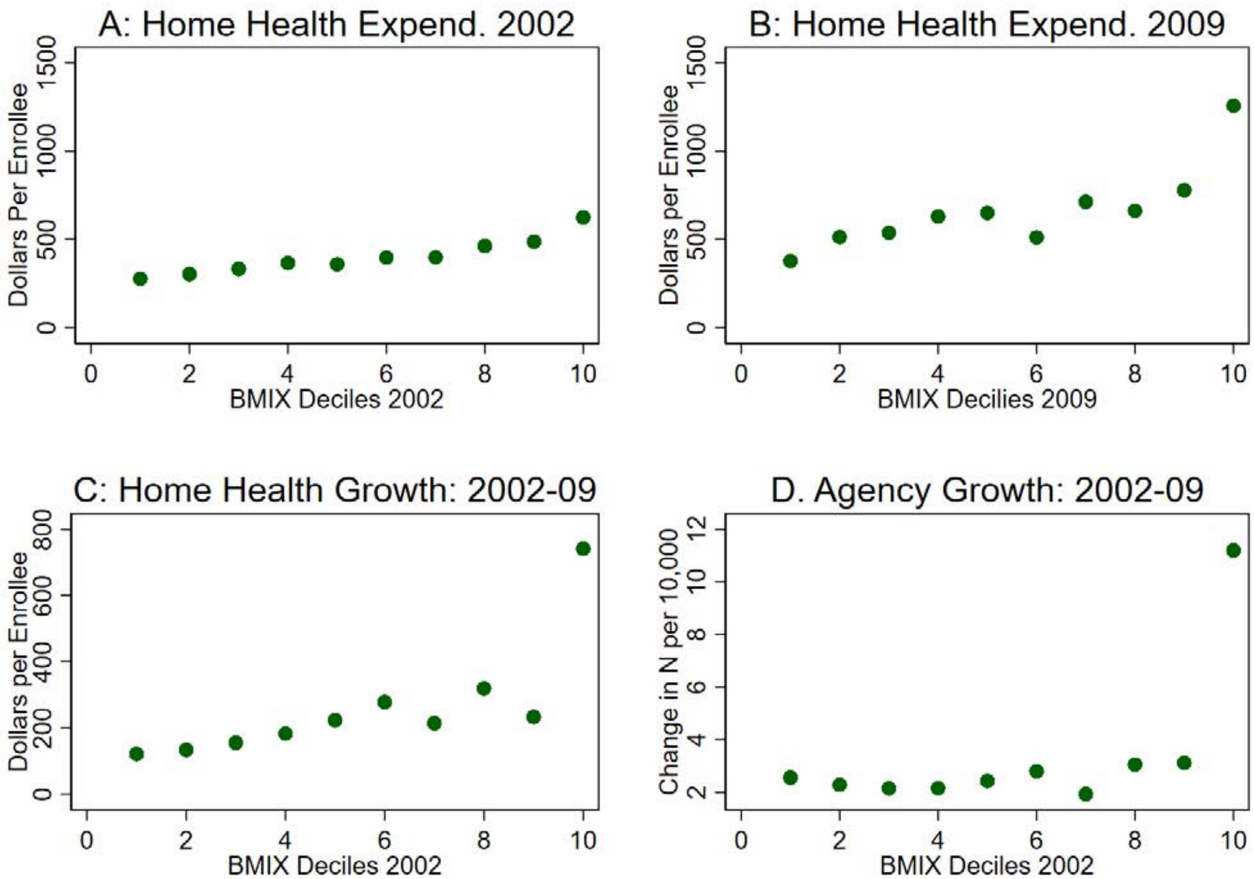


Figure 7: Home Health Care Expenditures in 2002 and 2009 and the Change in the Number of Home Health Agencies per 10,000 Medicare Enrollees, by Decile of the BMIX Index

Each decile corresponds to approximately 31 HRRs ranked in order of their BMIX index.

Table 1: Summary Statistics

	(1) Full sample	(2) DOJ Targeted HRRs (N=21)	(3) Other HRRs (N=285)
Home Health Expenditures	525.08	1152.27	478.87
	(332.6)	(512.0)	(261.5)
Home Health Exp.: 2002	378.84	671.18	357.30
	(166.4)	(192.1)	(142.4)
Home Health Exp.: 2009	642.00	1578.82	572.97
	(444.1)	(676.8)	(329.6)
Home Health Exp.: 2016	534.16	983.00	501.09
	(263.7)	(249.9)	(232.6)
BMIX	0.15	0.22	0.14
	(0.0749)	(0.115)	(0.0676)
No. of Agencies/10,000	3.20	8.03	2.85
	(3.119)	(7.210)	(2.186)
Network Density	0.34	0.30	0.34
	(0.236)	(0.243)	(0.235)
Betweenness Centrality	0.22	0.15	0.22
	(0.175)	(0.115)	(0.177)
Network Transitivity	0.54	0.48	0.54
	(0.234)	(0.230)	(0.234)
Mortality (per 1,000)	4.98	5.19	4.96
	(0.579)	(0.698)	(0.566)
FFS Medicare Population	93.59	135.71	90.49
	(87.69)	(154.2)	(79.81)
Observations	4590	315	4275

Measured at the HRR/Year level. Standard deviations in parentheses. FFS Denotes “Fee for Service.”

Table 2A: Regressions Explaining Home Health Expenditures

	OLS	OLS (Weighted)	OLS (2003-16)	Fixed Effect	Fixed Effect (Weighted)
BMIX*	173.2	164.1	86.42	31.59	66.91
	(6.59)	(5.67)	(3.20)	(1.79)	(2.27)
Network Density*	10.37	36.71	-53.59	-18.15	-36.29
	(0.77)	(2.27)	(-2.98)	(-2.47)	(-3.43)
Betweenness Centrality*	-34.87	-58.89	-41.15	-12.29	-19.11
	(-3.54)	(-4.07)	(-4.49)	(-3.71)	(-3.34)
Network Transitivity*	-58.43	-127.3	-60.21	-17.47	-42.80
	(-4.42)	(-4.64)	(-4.70)	(-3.89)	(-3.52)
Mortality (per 1,000)	203.7	165.9	213.1	25.17	58.52
	(8.03)	(6.76)	(8.17)	(1.32)	(2.57)
N of Agencies in 2002	1.795	0.115	1.509		
	(2.43)	(0.17)	(1.63)		
N Enrollees 2002 (1000)	-1.192	-0.512	-1.278		
	(-4.09)	(-2.29)	(-4.42)		
BMIX in 2002*			110.1		
			(4.21)		
Density in 2002*			64.54		
			(3.60)		
Betweenness in 2002*			25.98		
			(1.96)		
Transitivity in 2002*			15.66		
			(1.02)		
Constant	-607.5	-442.4	-357.3	250.8	60.55
	(-4.00)	(-3.07)	(-2.89)	(2.40)	(0.50)
Observations	4496	4496	4149	4496	4496
R^2	0.429	0.495	0.486	0.893	0.895

t statistics in parentheses. Year fixed-effects included in all regression models.
* Denotes z-score (standard error = 1). Clustered at the HRR level.

Table 2B: Regressions Explaining Home Health Care Expenditures (in Logs)

	OLS	OLS (Weighted)	OLS (2003-16)	Fixed Effect	Fixed Effect (Weighted)
BMIX*	0.308	0.273	0.159	0.0619	0.0972
	(10.42)	(9.66)	(4.59)	(4.08)	(6.69)
Network Density*	0.0491	0.0671	-0.0666	-0.0138	-0.0351
	(1.63)	(2.05)	(-2.47)	(-1.18)	(-2.38)
Betweenness Centrality*	-0.00555	-0.0432	-0.0134	-0.0108	-0.0161
	(-0.27)	(-1.72)	(-0.78)	(-2.22)	(-2.44)
Network Transitivity*	-0.0499	-0.140	-0.0489	-0.0206	-0.0415
	(-2.48)	(-3.99)	(-2.49)	(-3.71)	(-4.71)
Mortality (per 1,000)	0.387	0.311	0.381	0.0400	0.103
	(9.42)	(7.27)	(9.30)	(1.69)	(3.41)
Log N of Agencies 2002	0.0703	0.0135	0.217		
	(1.36)	(0.26)	(3.82)		
Log N Enrollees 2002	-0.161	-0.0820	-0.230		
	(-3.34)	(-1.75)	(-5.06)		
BMIX in 2002*			0.117		
			(3.27)		
Density in 2002*			0.183		
			(4.49)		
Betweenness in 2002*			0.0851		
			(3.39)		
Transitivity in 2002*			0.0722		
			(2.08)		
Constant	4.318	4.550	4.812	5.644	5.331
	(14.73)	(15.69)	(18.66)	(43.17)	(33.59)
Observations	4496	4496	4149	4496	4496
R^2	0.471	0.512	0.554	0.943	0.947

t statistics in parentheses. Year fixed effects included in all regressions.
* Denotes z-score (standard error = 1). Clustered at the HRR level.

Table 2C: Regressions Explaining the Number of Home Health Agencies (per 10,000)

	OLS	OLS: Log(Agencies)	OLS	Fixed Effect
BMIX*	15.39	0.450	20.12	14.64
	(2.39)	(16.37)	(2.32)	(1.56)
Network Density*	3.835	-0.419	-2.909	-7.657
	(1.57)	(-13.69)	(-1.48)	(-4.08)
Betweenness Centrality*	-1.490	-0.0747	-2.163	-2.347
	(-1.67)	(-4.52)	(-2.31)	(-3.26)
Network Transitivity*	-8.921	-0.0944	-9.806	-4.737
	(-3.55)	(-4.42)	(-3.79)	(-3.36)
Mortality (per 1,000)	-9.326	0.0702	-7.607	-7.685
	(-4.24)	(1.46)	(-3.48)	(-1.69)
N of Agencies in 2002	2.091	0.00853	2.382	
	(6.10)	(2.75)	(6.13)	
N Enrollees 2002 (1000)	-0.142	0.00197	-0.157	
	(-3.23)	(4.72)	(-3.40)	
BMIX in 2002*			-8.370	
			(-1.93)	
Density in 2002*			10.49	
			(4.84)	
Betweenness in 2002*			3.499	
			(2.66)	
Transitivity in 2002*			3.737	
			(2.22)	
Constant	45.16	2.026	29.85	67.29
	(3.01)	(7.38)	(2.67)	(2.95)
Observations	4496	4496	4149	4496
R^2	0.703	0.853	0.729	0.872

t statistics in parentheses. All regressions include year fixed-effects.
*Denotes z-score (standard error = 1). Clustered at the HRR level.

Table 2D: Regressions Explaining the Change (2002-2009) in Home Health Care Expenditures and Number of Agencies, and the Probability of Strike Force Designation

	Home Health	Home Health	Log Home Health	N of Agencies	Strike Force HRR
BMIX in 2002*	180.5	85.33	0.0883	1.479	0.0862
	(8.24)	(3.92)	(3.23)	(7.23)	(4.68)
Density in 2002*	-3.691	-9.711	0.0350	-0.461	0.0288
	(-0.16)	(-0.49)	(1.15)	(-2.19)	(1.52)
Betweenness in 2002*	16.36	0.605	0.0275	-0.100	0.0181
	(0.86)	(0.04)	(1.24)	(-0.57)	(1.13)
Transitivity in 2002*	10.29	11.04	0.0390	-0.212	-0.00718
	(0.44)	(0.54)	(1.49)	(-0.98)	(-0.37)
Mort. Change 2002-09	208.2	164.2	0.217	1.323	0.0141
	(3.95)	(3.53)	(3.65)	(2.69)	(0.32)
N of Agencies in 2002	2.973	2.079		0.0630	0.00386
	(2.62)	(2.08)		(5.96)	(4.05)
N Enrollees 2002 (1000)	-1.680	-1.231		-0.0216	-0.000897
	(-5.77)	(-4.74)		(-7.97)	(-3.66)
Home Health 2002		0.939			
		(9.39)			
Log N of Agencies 2002			0.156		
			(3.34)		
Log N Enrollees 2002			-0.182		
			(-4.80)		
Log Home Health 2002			-0.0128		
			(-0.26)		
Constant	469.4	67.83	0.990	4.490	0.0772
	(12.05)	(1.24)	(3.11)	(12.35)	(2.36)
Observations	300	300	300	300	300
r2	0.310	0.470	0.181	0.396	0.217

t statistics in parentheses. The Strike Force regression (final column) is a linear probability model, where the dependent variable is 1 if the region became a DOJ-targeted HRR.
* Denotes z-score (standard error = 1).

Table 3: Peer-Agency Associations of Log-Peer Home Health Spending with Log Ego Spending

	(1)	(2)	(3)	(4)
Average Linked Home Health Spending	-0.004 (.002)	0.003 (.003)	0.036*** (.003)	0.021** (.004)
Number of Peers	0.327*** (.002)	0.029*** (.003)	0.332*** (.002)	0.018*** (.004)
Average Linked Home * Number of Peers			0.036*** (.002)	0.009*** (.002)
Lagged HRR Isolate Home Health Spending	0.233*** (.004)	0.160*** (.003)	0.210*** (.004)	0.154*** (.004)
Lagged Ego Home Health Spending		0.608*** (.003)		0.607*** (.003)
R^2	0.536	0.738	0.538	0.738
Bayesian Information Criterion (BIC)	159,540	123,133	159,193	123,114

All models include HRR fixed-effects and agency (HHA) random-effects and are estimated on N = 126,749 agency-by-year observations involving 14,326 distinct agencies across the 306 HRRs. The R^2 measure is computed with the random-effects for agency being part of the error-term; this quantity is often referred to as marginal R^2 for a mixed-effect model. Smaller values of the BIC represent superior model fit. However, because the BIC increases with the sample-size, it only makes sense to make comparisons within models (1) and (3) and within models (2) and (4).

Appendix

A.1: Highest-Spending HRRs (2009 or 2010)

	HRR Name	Year	Home Health Expenditures
1.	*TX-McAllen*	*2009*	*3285.4*
2.	*FL-Miami*	*2009*	*3228.8*
3.	*TX-Harlingen*	*2010*	*2666.6*
4.	LA-Monroe	2009	2107.3
5.	*TX-Corpus Christi*	*2010*	*2029.1*
6.	LA-Alexandria	2010	1985.6
7.	*TX-Dallas*	*2009*	*1851.5*
8.	TX-Longview	2009	1691.3
9	*LA-Metairie*	*2010*	*1663.2*
10.	*IL-Chicago*	*2010*	*1623.2*
11.	*TX-Beaumont*	*2010*	*1604.6*
12.	*LA-Baton Rouge*	*2010*	*1592.4*
13.	TX-Wichita Falls	2009	1561.7
14.	*LA-New Orleans*	*2010*	*1554.3*
15.	*LA-Slidell*	*2009*	*1503.1*

Note: HRR is denoted in italics (and green) if it is designated by the DOJ as one of the strike force locations. Anecdotal evidence suggests that the adjacent DOJ-designated strike force teams are prosecuting cases even within the four regions in the top-15 not formally designated (Monroe LA, Alexandria, LA, Longview TX, and Wichita Falls TX). For example, one doctor from Alexandria LA was accused of nearly $5 million in falsified medical orders, prosecuted by the Gulf Coast DOJ law enforcement office (https://www.justice.gov/usao-edla/pr/gulf-coast-health-care-fraud-law-enforcement-action-results-charges-against-34). In another case, a West Monroe LA resident's case involving pharmaceutical health care fraud was prosecuted through the New Orleans field office (https://www.justice.gov/usao-sdms/pr/louisiana-man-sentenced-3-years-prison-conspiracy-commit-health-care-fraud)