



# The diffusion of health care fraud: A bipartite network analysis

A. James O'Malley<sup>a,b,\*</sup>, Thomas A. Bubolz<sup>b</sup>, Jonathan S. Skinner<sup>b,c,d</sup>

<sup>a</sup> Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, 1 Medical Center Drive, Lebanon, NH 03755, USA

<sup>b</sup> The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth, 1 Medical Center Drive, Lebanon, NH 03755, USA

<sup>c</sup> Department of Economics, Dartmouth College, Hanover, NH 03755, USA

<sup>d</sup> National Bureau of Economic Research, Cambridge, MA 02139, USA

## ARTICLE INFO

Handling Editor: Social Epidemiology Editorial Office

### Keywords:

Health care fraud  
Beneficiary-sharing  
Bipartite mixture network index  
Bipartite degree  
Diffusion  
Home health care  
Peer-effect  
Social network

## ABSTRACT

Many studies have examined the diffusion of health care innovation 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 (DOJ) anti-fraud “strike force” offices. We hypothesize that patient-sharing across home health care agencies (HHAs) provides a mechanism for the rapid diffusion of fraudulent strategies. We measure such activity using a novel bipartite mixture (or BMIX) network index, which captures patient sharing across multiple agencies and thus conveys more information about the diffusion process than conventional unipartite network measures. Using a complete population of fee-for-service Medicare claims data, we first find a remarkable increase in home health care activity between 2002 and 2009 in many regions targeted by the DOJ; average billing per Medicare enrollee in McAllen TX and Miami increased by \$2127 and \$2422 compared to just an average \$289 increase in other HRRs not targeted by the DOJ. Second, we establish that the HRR-level BMIX (but not other network measures) was a strong predictor of above-average home health care expenditures across HRRs. Third, within HRRs, agencies sharing more patients with other agencies were predicted to increase billing. 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 and contagion models more generally.

## 1. Introduction

Since the landmark study by James Coleman et al. (1966) of tetra-cycline, 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 through networks over time and across regions. Several studies have focused on the incentives for expanded (or reduced) Medicare fraud (Silverman and Skinner, 2004; Dafny, 2005; Leder-Luis, 2020; Eliason et al., 2021), which has been estimated to account for 8% of Medicare expenditures, or \$52 billion in 2017 (GAO, 2017). These studies, however, did not consider the potential role of networks in the diffusion process. Given the importance of networks in the dissemination of investment fraud (Baker, 2003; Nash et al., 2013), we hypothesize that networks play a key role in 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 \$3229 in 2009 per Medicare enrollee (in 2016 dollars; averaging over all fee-for-service Medicare enrollees, not just those receiving home health services). By contrast, in Los Angeles, home health billing (used interchangeably with expenditures) barely budged, from \$782 in 2002 to \$861 in 2009, a 10 percent increase. Largely in response to the rapid growth of home health billing 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.

In a pioneering study, Glaeser et al. (1996) suggested that the wide

\* Corresponding author. Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, 1 Medical Center Drive, Lebanon, NH 03755, USA.

E-mail addresses: [james.omalley@dartmouth.edu](mailto:james.omalley@dartmouth.edu) (A.J. O'Malley), [tom.bubolz@gmail.com](mailto:tom.bubolz@gmail.com) (T.A. Bubolz), [jonathan.skinner@dartmouth.edu](mailto:jonathan.skinner@dartmouth.edu) (J.S. Skinner).

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. The observed geographical variation in home health care expenditures is consistent with this approach; because Home Health Agencies (HHAs) are distinct organizations, our study of the spread of extreme Medicare billing from HHA to HHA is an organizational-level study of the dark side of social capital (Villalonga-Olives and Kawachi, 2017). 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., 2020), or variations across regions in social capital and physician professionalism (Skinner, 2011).

To better understand the remarkable growth of Medicare fraud during the 2000s in just a few regions of the U.S., we focus on the network structure of HHAs. We build on an economic model of criminal behavior (Becker; 1968; Sah, 1991) in which the extent of fraud depends on perceived profits and legal penalties, for example the introduction of criminal and civil legal proceedings against agency owners and physicians (Leder-Luis, 2020), or the imposition of pre-authorization policies and coverage denials (Eliaison et al., 2021; Howard and Desai, 2020). We expand on these models, however, to hypothesize that patient-sharing networks across home health agencies are central to the diffusion of fraud.

To capture the potential speed and extent of diffusion, 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 a greater number of patients shared between just two agencies. 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 homes networks of homes based on employees that are contracted by multiple of them, the travelling between them by such employees may be associated with the diffusion of COVID-19 outbreaks during the early days of the pandemic (Chen et al., 2020).

For background, we next describe the historical growth (and subsequent decline) in home health expenditures from 2002 to 2016 and then develop a theoretical framework to explain these changes and motivate the bipartite BMIX index, which is to our knowledge a new network index.

### 1.1. Spatiotemporal patterns of medicare home health care expenditures

The Medicare federal insurance program for people aged 65 and older in the U.S. 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 HHA 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).

**Home health fraud.** We define fraud as either knowingly billing for services with no benefit for (or even harm to) patients, or billing for services not provided. Only a fraction of fraudulent behavior is brought to trial, and innocent providers could also be falsely accused; thus true fraud is very difficult to measure. Instead, we proxy for fraud in a probabilistic sense by using “outlier” expenditure rates for otherwise similar patients that we interpret as reflecting a heightened probability of fraud, an approach that has been used in other studies to detect fraud (Shekhar et al., 2022). For example, one algorithm relying on outliers

identified 17 potentially fraudulent dental providers, of which 12 (71%) were recommended by auditors for further investigation (van Capelleveen et al., 2016). At the regional level, we use a different proxy for high rates of fraudulent activity: the presence of a local “strike-force” office of the DOJ and HHS devoted to detecting and prosecuting fraud, described below.

How does home health care fraud take place? It's difficult to generalize, but there are common patterns of fraudulent behavior, which includes agency owners providing home health services to Medicare beneficiaries which were not medically necessary and often were never provided. These efforts included kickbacks to physicians, patient recruiters and staffing groups to refer patients to their agency (HHS/DOJ, 2017, p. 21), or the sharing of patient IDs across networks of HHAs owned by organized criminal organizations (Meyers, 2017). (See Supplemental Appendix Section A.1 for more details.)

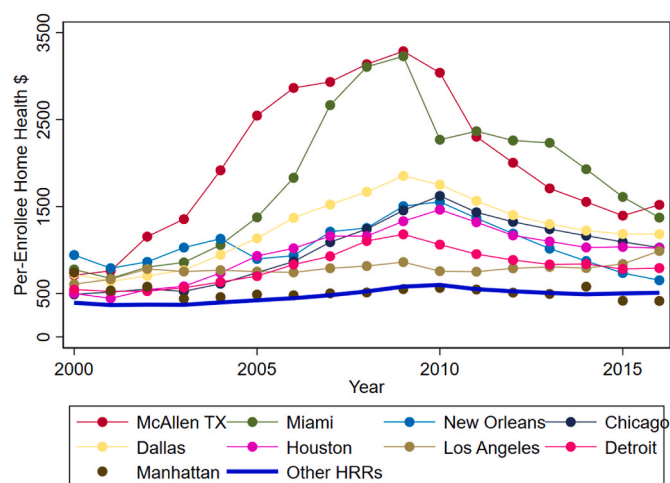
**Geographic variation in home health care expenditures 2002–16.** Home health care expenditures per Medicare fee-for-service enrollee are derived from age-sex-race-adjusted measures in the Dartmouth Atlas from 2002 to 16 for 306 hospital referral regions (HRRs) (Dartmouth Atlas, 2021). The comparisons 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.

While there are 306 HRRs in the U.S., we focus on those with documented evidence of fraudulent behavior; regions in which by 2016 the DOJ 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, we designated 21 regions deemed subject to strike force interest (See Supplemental Appendix Table A1 for a list of the strike force office locations and regions and Section A.2 for further discussion.). To provide visual clarity in our graphs, we focus on 9 of the larger regions.

Fig. 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 the network analyses in this paper begin in 2002 when the 100% fee-for-service data became available, we show in Fig. 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. 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 DOJ 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. 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 billing in either 2009 or 2010, 11 of them are in our designated targeted list of regions by the DOJ (Table A1); the remaining 4 are located in Texas and Louisiana within driving distance of strike-force HRRs.

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-billing HRRs in 2009 and 2010, but other Texas HRRs such as El Paso and Temple were



**Fig. 1.** Per-Enrollee Home Health Expenditures, 2000–2016, by Selected Region

Note: Selected regions from the 21 Department of Justice (DOJ) targeted HRRs 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.

much closer to the U.S. average. Similarly, Miami is an outlier even within Florida; the Fort Lauderdale HRR, adjacent to Miami, accounted for \$1175 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.

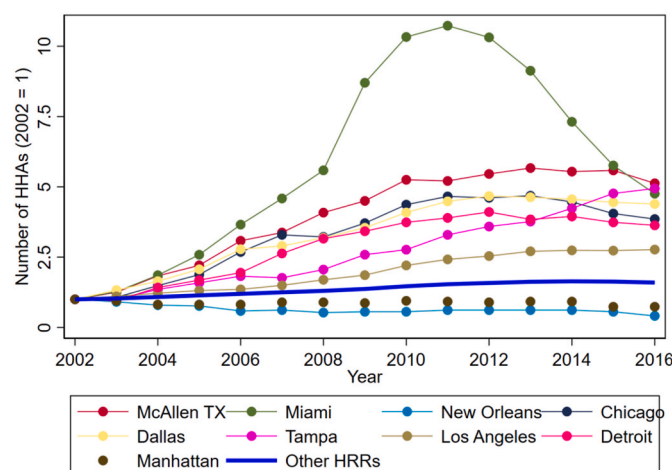
Fourth, there is a distinct rise and then decline in home health care expenditures, particularly for those targeted by the DOJ. The decline is likely to have been associated with two factors. The first is changes in policies enacted in response to potentially fraudulent activities, such as restrictions on outlier payments above the 60-day limit. Miami-Dade County accounted for nearly half of all U.S. home health outlier payments in 2009 (Benzio, 2010), so that when CMS sharply restricted outlier payments in 2010 (Kim and Norton, 2015), there was a particularly sharp decline in health care billing for the Miami HRR in 2010 (Fig. 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; Leder-Luis (2020) found that Qui Tam or “Whistleblower” provisions for Medicare and Medicaid fraud led to a \$6.80 specific deterrence effect per dollar of settlement.

A key feature of the expansion in home health care expenditures was an increase in the number of HHAs in areas identified by the DOJ. In Fig. 2, we show the number of HHAs 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. We next posit a theoretical model that links network factors involving market structure to a rapid diffusion in health care fraud.

## 1.2. Theoretical model: the role of patient referral networks

Here we outline a theory to motivate the hypothesis that a bipartite (beneficiary-HHA) network feature predicts the diffusion of fraudulent billing; this model is described in more detail in Supplemental Appendix Section A.3 and in O’Malley et al. (2021). We assume three types of HHAs: (a) those maximizing net social benefits arising from patient treatment and care; (b) those maximizing legal profits rather than social benefits (sometimes termed supplier-induced demand), and (c) those willing to risk conviction, fines, and imprisonment to make larger profits (Becker, 1968). In the short-term, the quantity of services and total



**Fig. 2.** The Number of Distinct Home Health Care Agencies (HHAs) Relative to 2002, by Selected Region

Note: Selected regions from the 21 DOJ-targeted HRRs 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 HHAs because of a decline in population after Hurricane Katrina in 2005.

billing for agencies described by (a) and (b) are not likely to change absent a dramatic increase in patients needing home health care or a fundamental change in prices or home health care technology; we do not observe such shifts in the Medicare program during the period of analysis. Indeed, for most regions in the U.S. we observe only gradual changes in utilization and billing, suggesting that in these regions, most agencies are types (a) and (b).

However, the model predicts a rapid expansion of type (c) agencies through fraudulent billing and patient sharing when (1) agencies are reassured that other agencies are behaving with similar practices and haven’t been caught, (2) knowledge about how to bill fraudulently is discovered through informal networks, and where (3) the likelihood of one agency learning fraudulent strategies from another is a positive but diminishing function of the number of patients shared between the two agencies. This latter condition means that a single patient from a type (c) agency shared with multiple (say 5) agencies predicts greater market fraud than if 5 patients from the (c) agency were shared with just one other agency. Conditions (1)–(3) imply that a bipartite measure of patient sharing would better distinguish between regions most amenable to the rapid diffusion of fraudulent activity, and those least amenable.

Traditional network measures are typically unipartite, meaning that the unit of analysis is (in our case) the HHA, with all information about specific patients lost. By contrast, a bipartite network has two distinct sets: One is the set of HHAs, the second the set of patients, and we measure the links between each patient and each HHA; thus we know (for example) whether a patient was shared exclusively between two HHAs, or whether that patient was shared across a wide set of HHAs. Furthermore, a bipartite measure captures patients whose care is solely provided by a given HHA, information invisible for unipartite network measures based solely on observed edges between HHAs. In sum, this bipartite approach provides valuable insights about the extent to which information may be shared across HHAs that would otherwise be lost in a unipartite network measure.

## 2. Methods and models

In this section, we develop a bipartite measure of exposure to potentially fraudulent agencies, describe the development of the unipartite regional networks and associated summary measures, describe the billing models involving the bipartite and unipartite network

measures, develop peer-effect models of diffusion in home health billing (or expenditures), and describe the empirical data and methodology for all analyses.

### 2.1. The BMIX index

Let  $B_{hi}$  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:

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

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  $S_i = \sum_{h=1}^H B_{hi} I(\sum_{k \neq i} B_{hk} = 0)$  is the total number of patients that receive support from agency  $i$  alone ("single-source patients"), 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 pairwise instances of 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 the number of single-source patients ( $S_i$ ) will aid its predictive power.

Further insight into the BMIX index arises by defining a patient attribute, denoted  $N_h$  for patient  $h$ , corresponding to the number of distinct agencies they received care from,  $N_h = \sum_{i=1}^n B_{hi}$ . Denote the number of patients with attribute value  $z$  by  $d_z = \sum_{h=1}^N I(N_h = z)$ . Because  $d_1 = \sum_{i=1}^n S_i$  and  $\sum_{i < k} A_{ik} = \sum_{z=2}^n \binom{z}{2} d_z = \frac{1}{2} \sum_{z=2}^n z(z-1) d_z$  it follows that:

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

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

The expression in (2) 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 contributed by a patient who encounters  $z$  agencies. As the number of their agency encounters increases, the impact a patient has on the BMIX index increases quadratically.

We are not aware of BMIX having been previously developed in the network literature (although it is related to market overlap measures, as in Aryal et al., 2020). The weights  $w_z$  are a mixture of patients that do ( $z > 1$ ) and do not ( $z = 1$ ) contribute to the network, making BMIX a combination of two forms of information. As noted in Section 1.2, knowledge of the number of patients shared between exactly two versus a wider set of agencies and the number of beneficiaries whose care is solely provided by a single agency is lost under the bipartite to unipartite projection. Low values of the BMIX index correspond to where patients remain with a single HHA for all their treatment, while larger values are consistent with jumps to multiple agencies, whether randomly or

because of explicit coordination among interlocking HHAs; the BMIX might be thought of as measuring the energy, pressure or heat (e.g., enthalpy) in a market.

To illustrate the calculation of BMIX, suppose that two HRRs each have 10 agencies, with 20 patients in total. In the first, 9 beneficiaries receive services from exactly 2 of the agencies and 11 receive services from just one agency. In the second, 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. See Supplemental Appendix Section A.4 for another worked example and more discussion of the BMIX.

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 all beneficiaries receive care from two or more agencies ( $d_1 = 0$ ) compared to the counterfactual that all beneficiaries receive care from a single agency.

### 2.2. Unipartite networks and measure development

To determine whether the BMIX index contains information beyond that captured in standard network measures, we constructed a beneficiary-sharing network for each HRR in each year from which standard network measures could be computed. The nodes are the HHAs physically located in the HRR and the existence of an edge between two agencies indicates that at least one patient received care from both agencies during a calendar year; see Supplemental Appendix Section A.5 for more details of the construction of these unipartite networks and the computation of the summary measures listed below. We also align the number of shared patients during the year with each such edge (the values of  $A_{ik}$  in Equation (1)) for computing the BMIX for the HRR. Besides the BMIX index measure, we construct the following three unipartite network measures by HRR and year chosen because they have a theoretical and empirical basis for predicting fraudulent activity (Aven, 2015; Ferrara et al., 2014; O'Malley et al., 2021):

- 1) Density, the fraction of potential connections or edges among nodes (or agencies).
- 2) Betweenness centralization, a measure of heterogeneity in the extent that each agency intersects the information flow in the network.
- 3) 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.

The construction of the unipartite HHA network, the above network measures, and the BMIX is a distinct data wrangling activity from the construction of the outcome cohort, outcomes and non-network predictors.

While the BMIX index has a readily interpretable scale from zero to one, the distributions of the others vary substantially with the number of nodes or agencies in the network. We scale all four network measures by their standard deviations to facilitate interpretation.

### 2.3. Regression models of billing

The Medicare home health claims from the Dartmouth Atlas are used to create HRR-year level per-enrollee expenditures in the fee-for-service



population. The primary outcome variable is the average price, inflation, and age-sex-race adjusted payment per patient for an HRR and year and BMIX from current or prior years is the key predictor. In regression models of per enrollee billing, we include the HRR/year mortality rates for the entire Medicare population as predictors to adjust for differences in the health needs of the population. We considered performing analyses to determine what patient medical conditions were associated with the growth in home health care and adding these as case-mix controls, but were concerned about the potential reverse causality associated with comorbidity data given evidence that physicians coded patients inappropriately with diseases such as diabetes in order to justify unneeded billings (Chuchmach and Ross, 2019).

Our measures of home health care networks and other measures of utilization count only home health visits by residents of an HRR that occur at a HHA in that HRR. Each measure is constructed by year for each of the 306 HRRs from 2002 to 2016, leading to 4590 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 HHAs per 1000 population relative to their 2002 frequency as a measure of profitability, and (b) whether the HRR was designated of interest to the DOJ, as noted above.

We use conventional cross-sectional time-series regression analysis using inflation-adjusted dollar and log-dollar billing per Medicare enrollee under multiple model specifications. We also consider whether our set of network measures in 2002 predict home health care utilization or number of HHAs (per Medicare enrollee) in 2003–16 conditional on contemporaneous measures of these same indices. That is, in predicting (e.g.) 2012 home health billing in an HRR, we include both the contemporaneous 2012 BMIX measure, and the 2002 BMIX measure. Where appropriate, we cluster by HRR.

Finally, we ask whether network measures (or other measures) in 2002 could have predicted the subsequent growth in home health expenditures between 2002 and 09. Because network density and other

network-level BMIX, we hypothesize that both the level of billing (outlying behavior) and an agencies degree or number of peers in their HRR network (reinforcement through multiple exposures) combine to impart influence. That is, being connected to more agencies will reinforce a willingness to bill more, given the same average billing, and exposure to multiple instances of high billing will have a greater impact than exposure to a single instance of high billing.

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 agencies by  $A_{jt}$  with  $m$ th off-diagonal element  $A_{jtmn}$  indicating whether agencies  $m$  and  $n$  of HRR  $j$  provided services to any of the same beneficiaries in year  $t$ . 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 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 billing for shared patients of agency  $i$  is limited only to services received outside the  $i$ th agency. The  $i$ th element of the product of  $W_{jt}$  and  $Y_{jt}$ ,  $\bar{Y}_{ijt} = (WY)_{ijt}$ , is the average billing of the agencies in HRR  $j$  with which agency  $i$  shares a network edge (O'Malley et al., 2020).

To account for factors unrelated to peer effects at the HRR level that may vary over time, we adjust for HRR-wide average billing of agencies that are isolated nodes (they have no edges with any other agencies) in a given year, denoted  $IsoBilling_{j(t-1)}$ . Because billing has a highly skewed distribution, we take the respective logs of ego agency billing, average peer agency billing, and HRR-wide average billing by isolate agencies. For all HHAs sharing at least one patient with another agency and at least one single-source patient, the general model of interest is:

where  $\beta_{0t}$  denotes year fixed-effects,  $\beta_1$  is the association of the ego's prior year billing with its current billing,  $\beta_2$  measures the association of isolate home health billing (e.g., for those not sharing patients) averaged across the HRR in the prior year with its current billing; this is designed

$$\log(Y_{ijt}) = \beta_{0t} + \beta_1 \log(Y_{ij(t-1)}) + \beta_2 \log(IsoBilling_{j(t-1)}) + \beta_3 \log(D_{ij(t-1)}) + \beta_4 \log((WY)_{ij(t-1)}) + \beta_5 \log(D_{ij(t-1)}) \log((WY)_{ij(t-1)}) + \theta_j + \varepsilon_{ijt} \quad (3)$$

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 covariates to adjust for different HRR market sizes. As an effect-directionality test, we consider the reverse – did average home health billing in 2002 predict the change over time in the BMIX index from 2002 to 2009? Because a greater range of regions is conducive to more precisely estimating the effects of region-level network and other predictors on billing, we only report the results of regression models estimated on all regions.

#### 2.4. Peer effects and models, and HHA peer association analyses

Peer-effects, also known as social influence or contagion, have been linked to the spread of fraudulent financial activities; e.g., Glaeser et al. (1996) has suggested that fraudulent financial strategies spread locally by learning from nearby criminals in a process consistent with peer-effects (also see Zenou, 2003). 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.

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 the

to capture HRR-specific trends in patient health needs. The three key coefficients are  $\beta_3$ , the extent to which the number of peers of the agency (their network degree) is predictive of their billing (in the following year),  $\beta_4$ , the extent to which the average billing across peer agencies is predictive of the agency's own billing, and  $\beta_5$ , the modification of the peer average billing association by the focal agency's network degree. Finally,  $\theta_j$  is a fixed-effect of the HRR to capture permanent differences in underlying health and other factors across HRRs and  $\varepsilon_{ijt}$  is a within-HRR error term. We refrained from using models with peer-predictors from the current time-period as these would spuriously inflate the peer-effect due to agencies who share patients having also billed for those patients.

We consider four basic variants of Equation (3); one that excludes the own-agency lagged billing (e.g., setting  $\beta_1 = 0$ ) and one that doesn't, one that assumes the interaction effect is null (e.g., setting  $\beta_5 = 0$ ) and one that doesn't. To the extent that own-agency lagged billing 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 (the tendency of HHAs with similar home health care billing to subsequently share patients).

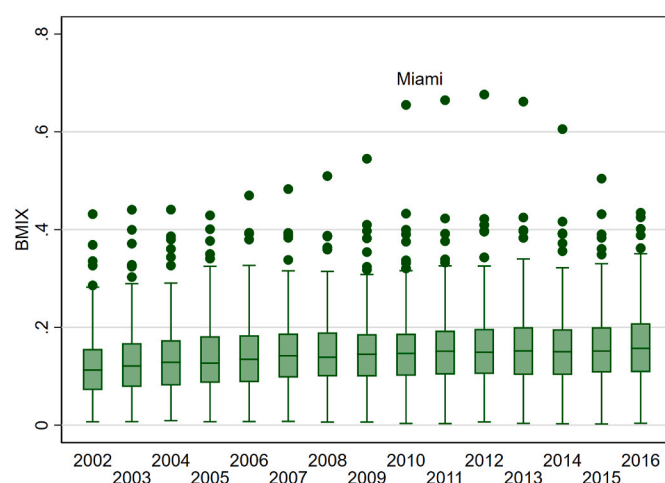
To guard against endogeneity from agencies that shared beneficiaries having their billing measure affected by the same patients, we restrict billing to the sample of beneficiaries that only receive care at a single agency (“single source” or degree 1 beneficiaries in the bipartite patient-HHA network). Such patients are the vast majority, reflecting

that the standard care is for patients to receive home health care from a single agency. The estimated peer-agency regression coefficient is thus informed only by patients that did not contribute to the formation of the network; the estimated coefficient 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 pre-standardized value of billing was positive (see [Supplemental Appendix Section A.6](#) for interpretation of the interaction effect coefficients). We treated each ego agency as a random effect 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.

### 3. Results

[Table 1](#) presents summary statistics on home health care expenditures, for specific years 2002, 2009, and 2016, and network measures. We find a remarkable amount of regional variation across HRRs in home health billing; the coefficient of variation rose from 0.44 in 2002 to 0.69 in 2009, before dropping to 0.49 in 2016. We further considered these summary statistics for DOJ-targeted HRRs, and those not targeted. For network measures, there is wide variability in the average value of the indices across regions, but the differences between the 21 DOJ-targeted and non-targeted 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 HHAs (8 compared to 2.8 per 10,000 Medicare beneficiaries) and home health expenditures (\$1152 compared to \$479) in the DOJ-targeted HRRs compared to the non-DOJ regions.

[Fig. 3](#) shows graphically the distribution of BMIX by year using a box-and-whisker graph; it exhibits wide variability around a mean of 0.15 with patient sharing rising through to the peak expenditures of 2009–10. Miami is a consistent extreme outlier (labeled); Fort Lauderdale, Las



**Fig. 3.** 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 (0.65), Las Vegas (0.43), Fort Lauderdale (0.40), Houston (0.39), and Los Angeles (0.37).

Vegas, Houston, and Los Angeles also had high rates of BMIX.

#### 3.1. The structure of patient-sharing BMIX networks

[Fig. 4](#) shows the Miami HRR networks in 2002 and 2009, while [Fig. 5](#) displays the patient-sharing network in Seattle (a low-growth region) during the same years. The 2009 Miami network 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 billing measures can be seen by sorting HRRs into deciles by their BMIX measure, either in 2002 or in 2009. In [Fig. 6](#), 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 top BMIX decile. The 2002 BMIX index predicts the subsequent growth in home health billing between 2002 and 09 (Panel C) and the corresponding growth in the number of HHAs 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.

#### 3.2. Billing regressions

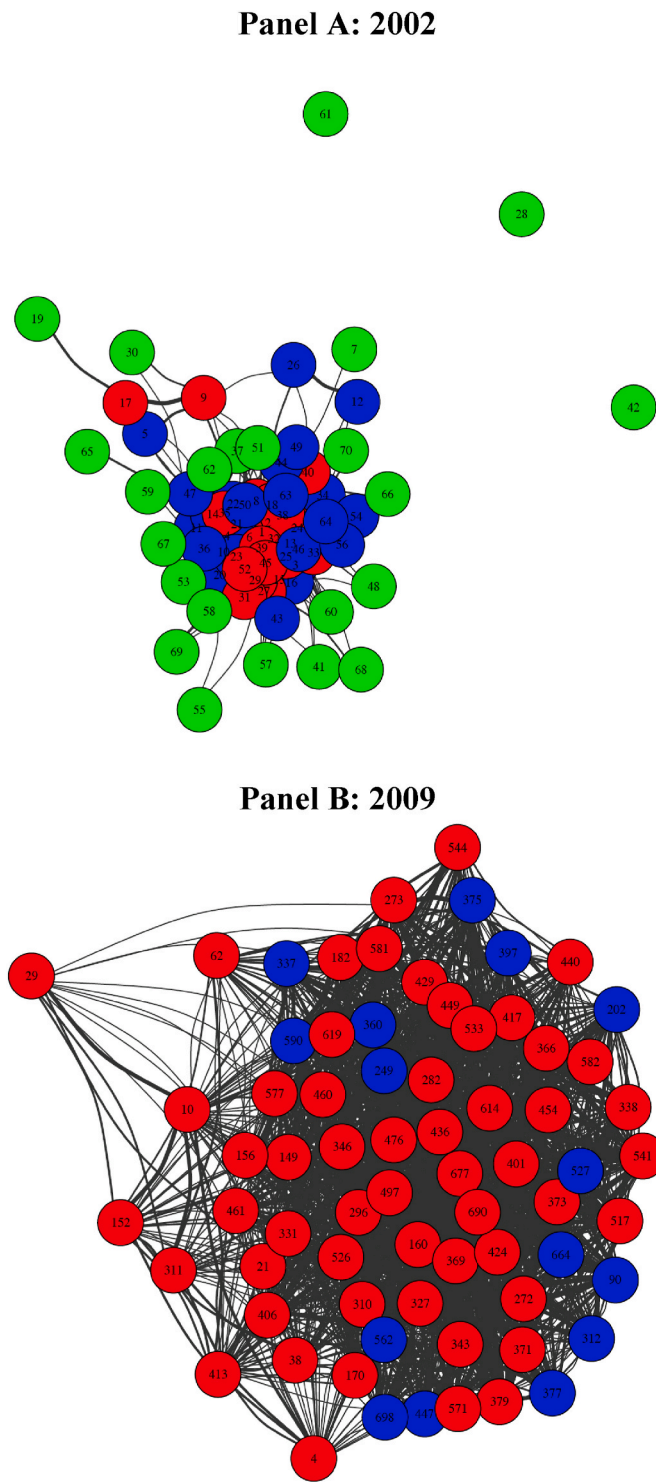
In the regression models in [Table 2A](#), home health care expenditure is regressed on multiple predictors, including the network measures. As noted above, each of the four network measures (starred) has a standard deviation of 1.0 and a mean of zero; the interpretation of each of these coefficients 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 level of mortality and are clustered by HRR. In the first column, a one-standard-deviation increase in BMIX is predicted to increase home health care billing by \$173, or 33 percent of

**Table 1**  
Summary statistics.

	(1) Full sample	(2) DOJ Targeted HRRs (N = 21)	(3) Other HRRs (N = 285)
Home Health Expenditures	525.08 (332.6)	1152.27 (512.0)	478.87 (261.5)
Home Health Exp.: 2002	378.84 (166.4)	671.18 (192.1)	357.30 (142.4)
Home Health Exp.: 2009	642.00 (444.1)	1578.82 (676.8)	572.97 (329.6)
Home Health Exp.: 2016	534.16 (263.7)	983.00 (249.9)	501.09 (232.6)
BMIX	0.15 (0.0749)	0.22 (0.115)	0.14 (0.0676)
No. of Agencies/ 10,000	3.20 (3.119)	8.03 (7.210)	2.85 (2.186)
Network Density	0.34 (0.236)	0.30 (0.243)	0.34 (0.235)
Betweenness Centrality	0.22 (0.175)	0.15 (0.115)	0.22 (0.177)
Network Transitivity	0.54 (0.234)	0.48 (0.230)	0.54 (0.234)
Mortality (per 1000)	4.98 (0.579)	5.19 (0.698)	4.96 (0.566)
FFS Medicare Population	93.59 (87.69)	135.71 (154.2)	90.49 (79.81)
Observations	4590	315	4275

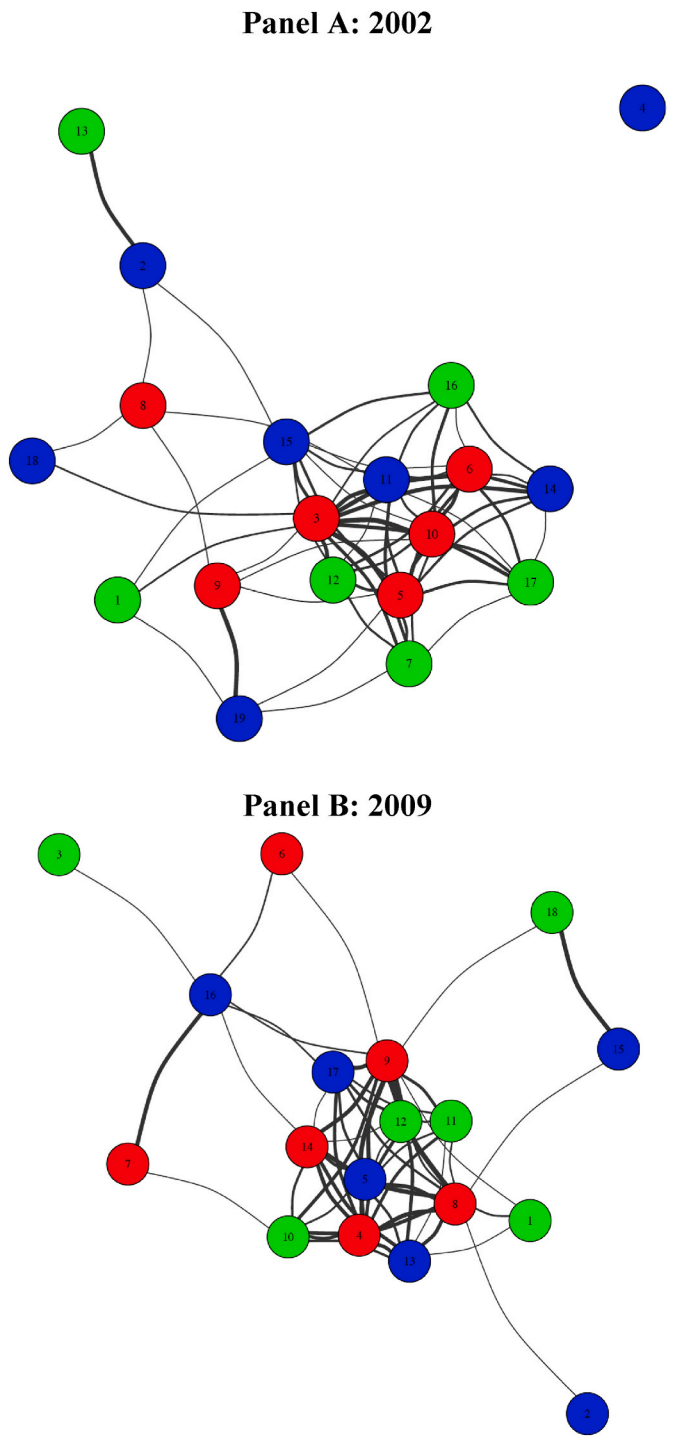
Measured at the HRR/Year level. Standard deviations in parentheses. FFS Denotes “Fee for Service.”



**Fig. 4.** 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.

average home health billing; other network measure coefficients are smaller in magnitude or negative.

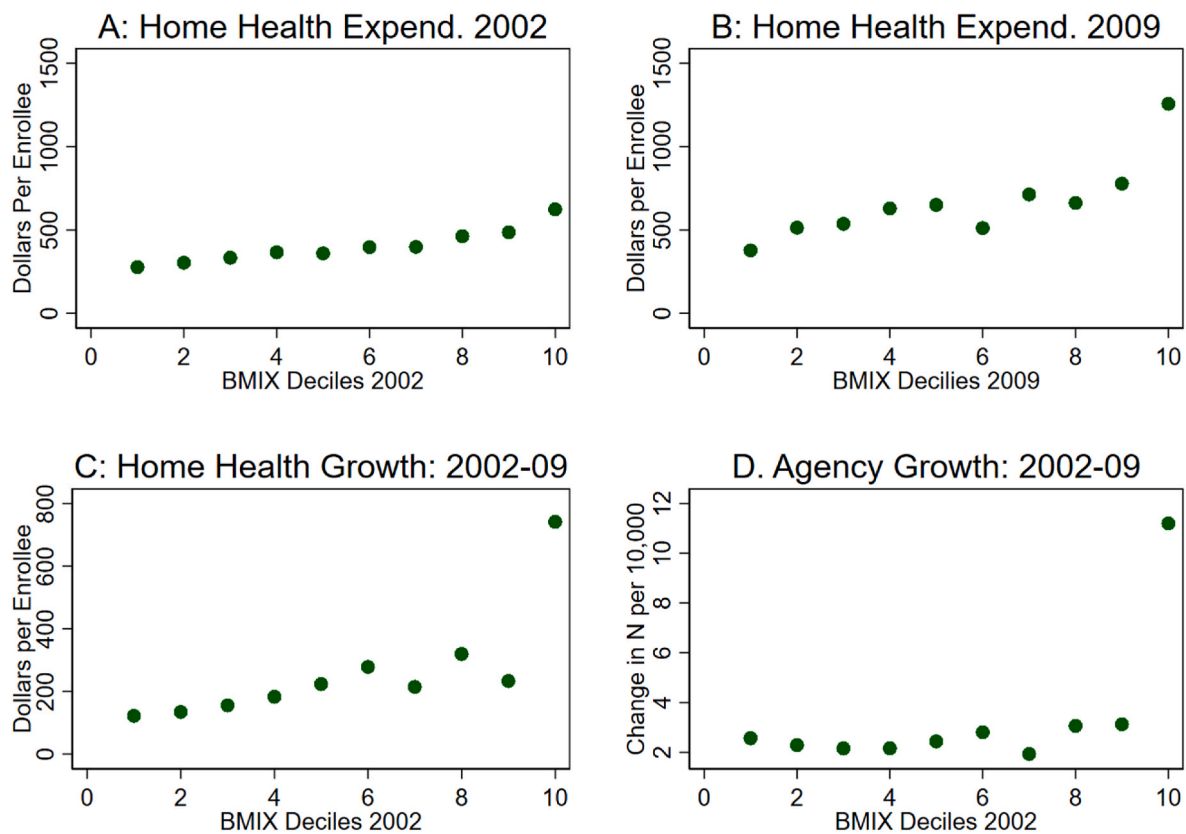
One can construct theories for why the other network measures would exhibit negative rather than positive associations. For example, higher transitivity could be an indication of greater coordination of unnecessary beneficiary sharing among “like” agencies (perhaps those



**Fig. 5.** 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.

owned centrally) within a region. Greater centralization might be 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.

A higher mortality rate is associated with higher home health billing; a one-standard deviation increase is predicted to increase home health billing by \$118 (the coefficient, \$203.7, times the standard deviation, from Table 1); weighted regression coefficients are similar. The Column 4 model, for years 2003–16, includes both contemporaneous and 2002 levels of (standardized) network measures. Once again, BMIX enters



**Fig. 6.** Home Health Care Expenditures in 2002 and 2009 and the Change in the Number of HHAs 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 2A**

Regressions explaining home health expenditures.

	OLS	OLS (Weighted)	OLS (Log)	OLS (2003–16)	Fixed Effect	Fixed Effect (Log)
BMIX*	173.2 (6.59)	164.1 (5.67)	0.308 (10.42)	86.42 (3.20)	31.59 (1.79)	0.0619 (4.08)
Network Density*	10.37 (0.77)	36.71 (2.27)	0.0491 (1.63)	-53.59 (-2.98)	-18.15 (-2.47)	-0.0138 (-1.18)
Betweenness Centrality*	-34.87 (-3.54)	-58.89 (-4.07)	-0.00555 (-0.27)	-41.15 (-4.49)	-12.29 (-3.71)	-0.0108 (-2.22)
Network Transitivity*	-58.43 (-4.42)	-127.3 (-4.64)	-0.0499 (-2.48)	-60.21 (-4.70)	-17.47 (-3.89)	-0.0206 (-3.71)
Mortality (per 1000)	203.7 (8.03)	165.9 (6.76)	0.387 (9.42)	213.1 (8.17)	25.17 (1.32)	0.0400 (1.69)
N of Agencies in 2002	1.795 (2.43)	0.115 (0.17)	0.0703 (1.36)	1.509 (1.63)		
N Enrollees 2002 (1000)	-1.192 (-4.09)	-0.512 (-2.29)	-0.161 (-3.34)	-1.278 (-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 (-4.00)	-442.4 (-3.07)	4.318 (14.73)	-357.3 (-2.89)	250.8 (2.40)	5.644 (43.17)
Observations	4496	4496	4496	4149	4496	4496
R <sup>2</sup>	0.429	0.495	0.471	0.486	0.893	0.943

*t* statistics in parentheses. All regressions include year fixed-effects.

\* Denotes z-score (standard deviation = 1). Clustered at the HRR level.

significantly, with a slightly larger coefficient for the 2002 value relative to the contemporaneous BMIX measure; their combined impact, corresponding to 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 (column 5), the association between BMIX and billing is much diminished, suggesting that billing does not track year-



**Table 2B**

Regressions explaining the change (2002–2009) in home health care expenditures and the probability of DOJ targeted HRRs.

	Home Health	Home Health	Log Home Health	DOJ Targeted HRRs
BMIX in 2002*	180.5 (8.24)	85.33 (3.92)	0.0883 (3.23)	0.0862 (4.68)
Density in 2002*	−3.691 (−0.16)	−9.711 (−0.49)	0.0350 (1.15)	0.0288 (1.52)
Betweenness in 2002*	16.36 (0.86)	0.605 (0.04)	0.0275 (1.24)	0.0181 (1.13)
Transitivity in 2002*	10.29 (0.44)	11.04 (0.54)	0.0390 (1.49)	−0.00718 (−0.37)
Mort. Change 2002–09	208.2 (3.95)	164.2 (3.53)	0.217 (3.65)	0.0141 (0.32)
N of Agencies in 2002	2.973 (2.62)	2.079 (2.08)		0.00386 (4.05)
N Enrollees 2002 (1000)	−1.680 (−5.77)	−1.231 (−4.74)		−0.000897 (−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 (12.05)	67.83 (1.24)	0.990 (3.11)	0.0772 (2.36)
Observations	300	300	300	300
r <sup>2</sup>	0.310	0.470	0.181	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 deviation = 1).

to-year with BMIX (or any of the other variables); we consider longer-term growth in home health care billing below in Table 2B. The results for the corresponding regressions of log of home health care billing tell a similar story with a one-standard deviation rise in BMIX predicted to increase log home health billing by nearly one third (column 3), with much smaller predicted effects in the HRR fixed effects model (column 6); however, the estimated coefficient of BMIX (0.0619) remains statistically significant (z-score >4). (See Supplemental Appendix Section A.7 and Tables A2 and A3 for sensitivity analysis using log billing regression models and the number of HHAs as dependent variables.)

Table 2B provides estimates of changes in home health expenditure between 2002 and 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 billing from 2002 is included in the regression (column 2). With lagged home health billing changes (column 3), the coefficient on BMIX is 0.088 (or 20 percent of the log change) with the 2002 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 billing in 2002 shows a negative (and barely significant) correlation. Finally, Table 2B shows in a linear probability model 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.

### 3.3. Peer association analyses

The results of the peer-associated analyses are based on 1.54 million shared home health care patients across HHAs in the 306 HRRs

**Table 3**

Peer-agency associations of log-peer home health billing with log ego billing.

	Model (1)	Model (2)	Model (3)	Model (4)
Average Linked Home Health Billing	−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 *			0.036*** (.002)	0.009*** (.002)
Number of Peers				
Lagged HRR Isolate Home Health Billing	0.233*** (.004)	0.160*** (.003)	0.210*** (.004)	0.154*** (.004)
Lagged Ego Home Health Billing		0.608*** (.003)		0.607*** (.003)
R <sup>2</sup>	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 home health care 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<sup>2</sup> measure is computed with the random-effects for agency being part of the error-term; this quantity is often referred to as marginal R<sup>2</sup> for a mixed-effect model. We also use the Bayesian Information Criterion (BIC) to compare the fitted models. 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).

(Table 3). We find the specifications with the interaction between average peer billing and number of peer agencies (i.e., degree) included yield superior fits than when this is excluded; the marginal R<sup>2</sup> (the agency random effects are in the error-term as opposed to the model fit) demonstrates an advantage of including the interaction term.

In the absence of the interaction, the dominant network-related node-level predictor is lagged log degree; in the model without lagged ego billing (e.g., the lagged dependent variable) the estimated coefficient is 0.327 (standard error 0.002) and in the model with lagged ego billing the estimated coefficient is 0.029 (standard error 0.003). In both models, lagged log peer average billing is statistically non-significant. We believe that these 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 is likely biased downward because the billing measure does not capture the dynamics by which peer associations in past years are already reflected in year  $t - 1$  ego billing measures.

With the addition of the interaction between average home health billing and network degree (number of peer agencies), the overall impact of lagged log peer average billing amplifies; in the model without lagged ego billing the main and interaction estimated coefficients are 0.036 (0.003) and 0.036 (0.002), respectively, and in the model with lagged ego billing they are 0.021 (0.004) and 0.009 (0.002). Considering the interaction terms in Column 4, the model predicts that when average lagged billing 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 billing by the ego is 0.033; the corresponding estimate for a one-standard-deviation increase in logged average billing at the 90th percentile for logged number of peers is 0.030. Finally, the prediction associated with a simultaneous increase in logged average billing and logged average number of peers from their means to their 90th percentiles is 0.065, a combination of results that parallels those for the diffusion of a medical procedure in O'Malley et al. (2020).

In sum, we have established that even *within* HRRs, HHAs sharing patients with a greater number of other agencies or with high-billing other agencies were more likely to increase patient expenditures in the following year. We also explored extending the model in Equation (3) 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.

#### 4. Conclusions

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, 2015) and newly developed cancer drugs (Agha and Molitor, 2018); however, much less is known about the network-based diffusion of ineffective or potentially harmful use of potentially fraudulent health care (Villalonga-Olives and Kawachi, 2017). In this paper, we have studied a rapid increase in billing for Medicare home health care expenditures in some 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. Instead, they appear largely the consequence of widespread fraudulent behavior which in turn attracted specific DOJ strike-force offices located in areas with rapid increases in fraudulent behavior.

Guided by a theoretical model of fraudulent billing in which the potential gains from such activity outweigh penalties of legal convictions, we developed a novel bipartite mixture network index, the BMIX. The crucial information-based features of the BMIX are that, unlike commonly used unipartite network measures, it retains knowledge of the number of patients shared between exactly two versus a wider set of agencies, as well as reflecting the fraction of beneficiaries whose care is solely provided by a single agency.

The BMIX index varied widely across regions and was strongly associated with per-enrollee home health care expenditures. Commonly used unipartite network measures such as density, betweenness-centralization, and transitivity were much less predictive of this rapid increase in home health billing. Notably the 2002 BMIX, measured at the outset of the sharp rise in home health care billing, was predictive of the growth in *subsequent* home health care expenditures for the period 2003–09 and was a strong predictor of the subsequent growth in the number of HHAs, and whether the region would attract a DOJ strike force office. The reverse did not hold; higher 2002 billing predicted slightly lower BMIX growth. Finally, we found evidence of peer associations *within* HRRs; HHAs sharing patients with multiple high-billing agencies were more likely to experience higher expenditures in the following year. The coefficient estimates are consistent with provider communication that, similar to Barnett et al. (2012) and O'Malley et al. (2020), occurs through the sharing of patients across agencies. In sum, our results suggest an important role for market bipartite network structure in the diffusion of fraudulent behavior.

Our study focuses on what is now a historical period of particularly high rates of home health billing, but new approaches to Medicare fraud have continued to undergo Darwinian evolution with novel schemes. For example, in July 2022 the Department of Justice brought charges of \$1.2 billion in fraudulent Medicare billing arising from a variety of approaches, including one in which “telemedicine companies found medical professionals” to prescribe “expensive genetic tests and durable medical equipment regardless of whether the patients needed them ...” (DOJ, 2022). We hypothesize that the use of a bipartite network structure for these schemes – or perhaps for schemes not yet discovered – can provide valuable predictive signals to federal agencies, because the sharing of patients (or patient identification) is a common feature of fraudulent medical billing. Identifying such activities early allows targeted pre-emptive auditing, which has been shown to provide large savings arising from deterrent effects among all providers, and not just those directly accused (Leder-Luis, 2020; Shi, 2022). That patient encounter data already exists, and is available with just a few months lag, makes the calculation of such bipartite network measures an attainable task at modest cost.

A second advantage of our network measures is that those committing fraud will have little idea of the extent to which the network makes them stand out because any one agency sees only what they do, not what other HHAs are doing. Therefore, networks may provide a covert monitor of provider behavior, at the level of an entire industry, across time. This may be a particularly helpful device to capturing general changes in patterns of care in close to real-time, and further allows us to measure how industry patterns change and which agencies or practices stand out.

We acknowledge three limitations of the analysis. The first is that we cannot measure fraud directly because agencies are understandably reticent about their potentially illegal behavior. Because those successfully charged are only the tip of the iceberg, most fraudulent activities are difficult to detect; we recognize that some of the three-fold increase in home health billing in Miami between 2002 and 2009 could have been legitimate. An organization with access to a database of confirmed cases (e.g., CMS) does not face this limitation; methods developed here could help identify areas where outbreaks of fraud appear the most likely.

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 in these nor in the peer-effect analyses (O'Malley et al., 2014). The BMIX index is not likely to be capturing unmeasured health effects – it is uncorrelated with mortality – but patient sharing patterns could be symptomatic of past or current fraudulent activity which in turn lays the groundwork for future fraudulent behavior. For example, Florida may experience permanently higher levels of fraud because of “corporate practice of medicine” laws that allow for HHAs to be owned by entrepreneurial non-clinicians (Health Law Firm Blog, 2012). State-level regulations also influence the ability of entrepreneurial physicians to form and profit from their own businesses (Welk, 2003).

Third, we cannot guarantee that the BMIX measure will continue to predict Medicare fraud as it evolves into schemes involving telemedicine, for example. While we believe that any rapid and clinically unjustified increase in billing will involve multiple providers relying on the same patient identification numbers (and thus triggering an elevated BMIX), we acknowledge that further research is necessary using more recent data.

Despite these limitations, the BMIX index has shown promise in predicting future excessive billing behavior for HHAs, suggesting its value for machine-learning approaches to unearthing Medicare fraud in HHAs and elsewhere (Bauder et al., 2017). 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. We also recognize that other bipartite measures (e.g., Opsahl, 2013; Opsahl et al., 2010) which are distinct from BMIX may also predict diffusion and should be considered in future research.

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). 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 in just 4 states: Pennsylvania, Maryland, Ohio, and Kentucky, a finding consistent with networks of interventional cardiologists within and across hospital systems. 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.

## Data availability

The data that has been used is confidential.

## Acknowledgements

We received funding from The National Institute on Aging for financial support (P01AG019783 and U01AG046830). We are grateful to Julie Bynum, Maxim Engers, Ambar La Forgia, Stephen Lee, Bruce Sacerdote, Owen Skinner, seminar participants at Dartmouth and Columbia, and especially Jetson Leder-Luis for valuable suggestions.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.socscimed.2023.115927>.

## References

- Agha, Leila, Molitor, David, 2018. The local influence of pioneer investigators on technology adoption: evidence from new cancer drugs. *Rev. Econ. Stat.* 100 (1), 29–44.
- Aryal, Gaurab, Campbell, Dennis J., Ciliberto, Federico, Khmelnitskaya, Ekaterina, 2020. Common-subcontracting and Multimarket Contact in the Airline Industry. University of Virginia Working Paper.
- Atlas, Dartmouth, 2021. The Dartmouth Atlas of health care. [www.dartmouthatlas.org](http://www.dartmouthatlas.org) (Accessed 10 December 2021).
- Aven, Brandy L., 2015. The paradox of corrupt networks: an analysis of organizational crime at Enron. *Organ. Sci.* 26 (4), 980–996.
- Baker, Wayne E., Faulkner, Robert R., 2003. Diffusion of fraud: intermediate economic crime and investor dynamics. *Criminology* 41 (4), 1173–1206.
- Barnett, Michael L., Christakis, Nicholas A., O'Malley, A. James, Onnela, Jukka-Pekka, Keating, Nancy L., Landon, Bruce E., 2012. Physician patient-sharing networks and the cost and intensity of care in US hospitals. *Med. Care* 50, 152–160.
- Bauder, Richard, Khoshgoftaar, Taghi M., Naeem, Seliya, 2017. A survey on the state of healthcare upcoding fraud analysis and detection. *Health Serv. Outcome Res. Methodol.* 17 (1), 31–55.
- Becker, Gary S., 1968. Crime and punishment: an economic approach. In: Becker, Gary S., Landes, William M. (Eds.), *The Economic Dimensions of Crime*. Palgrave Macmillan, London, pp. 13–68.
- Benzio, Brooke, 2010. Fee-for-Disservice: Medicare fraud in the home healthcare industry. *Annals of Health Law* Volume 19 Issue 1 (Article 41).
- Chen, M. Keith, Elisa, F., Chevalier, Judith A., 2020. Long. "Nursing Home Staff Networks and COVID-19." No. 27608. National Bureau of Economic Research.
- Chetty, Raj, Stepner, Michael, Abraham, Sarah, Lin, Shelby, Scuderi, Benjamin, Turner, Nicholas, Bergeron, Augustin, Cutler, David, 2016. The association between income and life expectancy in the United States, 2001–2014. *JAMA* 315 (16), 1750–1766.
- Chuchmach, Megan, Ross, Brian, <https://abcnews.go.com/Blotter/undercover-82-year-grandma-catches-medicare-fraud-tape/story?id=15818462>.
- Coleman, James, Katz, Elihu, Menzel, Herbert, 1966. *Medical Innovation: A Diffusion Study* Indianapolis. The Bobbs-Merrill Company, Inc.
- Cutler, David, Skinner, Jonathan S., Stern, Ariel Dora, Wennberg, David, 2019. "Physician beliefs and patient preferences: a new look at regional variation in health care billing. *Am. Econ. J. Econ. Pol.* 11 (1), 192–221.
- Dafny, Leemore S., 2005. How do hospitals respond to price changes? *Am. Econ. Rev.* 95 (5), 1525–1547.
- Dimmock, Stephen G., Gerken, William C., Graham, Nathaniel P., 2018. Is fraud contagious? Coworker influence on misconduct by financial advisors. *J. Finance* 73 (3), 1417–1450.
- DOJ (U.S. Department of Justice), 2022. Justice Department Charges Dozens for \$1.2 Billion in Health Care Fraud. Available at: <https://www.justice.gov/opa/pr/justice-department-charges-dozens-12-billion-health-care-fraud>.
- Eliason, Paul J., League, Riley J., Leder-Luis, Jetson, McDevitt, Ryan C., Roberts, James W., 2021. Ambulance Taxis: the Impact of Regulation and Litigation on Health Care Fraud. National Bureau of Economic Research. No. w29491.
- Ferrara, Emilio, De Meo, Pasquale, Catanese, Salvatore, Fiumara, Giacomo, 2014. Detecting criminal organizations in mobile phone networks. *Expert Syst. Appl.* 41 (13), 5733–5750.
- GAO, 2017. Medicare and Medicaid. CMS Needs to Fully Align its Antifraud Efforts with the Fraud Risk Framework. GAO-18-88 (Washington, D.C.: Dec. 5).
- Glaeser, Edward L., Sacerdote, Bruce, Scheinkman, Jose A., 1996. Crime and social interactions. *Q. J. Econ.* 111 (2), 507–548.
- Gottlieb, Daniel, J., Zhou, Weiping, Song, Yunjie, Andrews, Kathryn, Skinner, Jonathan, Sutherland, Jason, M., 2010. Prices don't drive regional Medicare billing variations. *Health Aff* 29 (3), 537–543.
- Health Law Firm Blog, 2012. The corporate practice of medicine in Florida. <http://healthlawfirm.blogspot.com/2012/05/corporate-practice-of-medicine-in.html>.
- HHS/DOJ, 2017. 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. <https://oig.hhs.gov/publications/docs/hcfac/FY2016-hcfac.pdf>.
- Howard, David H., Desai, Nihar R., 2020. US False Claims Act investigations of unnecessary percutaneous coronary interventions. *JAMA Intern. Med.* 180 (11), 1534–1536.
- Imperato, Gabriel, 2017. An Overview of Pervasive Allegations in Hospice and Home Health Fraud Cases. Broad and Cassel, Fort Lauderdale, FL. [https://www.hcca-info.org/Portals/0/PDFs/Resources/Conference\\_Handouts/Regional\\_Conference/2017/orlando/ImperatoHandout2.docx](https://www.hcca-info.org/Portals/0/PDFs/Resources/Conference_Handouts/Regional_Conference/2017/orlando/ImperatoHandout2.docx).
- Jencks, Stephen F., Huff, Edwin D., Cuerdon, Timothy, 2003. Change in the quality of care delivered to Medicare beneficiaries, 1998–1999 to 2000–2001. *JAMA* 289 (3), 305–312.
- Kim, Hyunjee, Norton, Edward C., 2015. Effects of the ten percent cap in Medicare home health care on treatment intensity and patient discharge status. *Health Serv. Res.* 50 (5), 1606–1627.
- Leder-Luis, Jetson, 2020. Working Paper Whistleblowers, Private Enforcement, and Medicare Fraud. Boston University. <https://sites.bu.edu/jetson/files/2020/07/Fals-e-Claims-Act-Paper.pdf>.
- Meyers, Travis J., 2017. Examining the network components of a Medicare fraud scheme: the Mirzoyan-Terdjanian organization. *Crime Law Soc. Change* 68 (1), 251–279.
- Nash, Rebecca, Bouchard, Martin, Malm Aili, 2013. Investing in people: the role of social networks in the diffusion of a large-scale fraud. *Soc. Network.* 35 (4), 686–698.
- O'Malley, A. James, Elwert, Felix, Rosenquist, J. Niels, Zaslavsky, Alan M., Christakis, Nicholas A., 2014. Estimating peer effects in longitudinal dyadic data using instrumental variables. *Biometrics* 70 (3), 506–515.
- O'Malley, A. James, Moen, Erika L., Bynum, Julie P.W., Austin, Andrea M., Skinner, Jonathan S., 2020. Modeling peer effect modification by network strength: the diffusion of implantable cardioverter defibrillators in the US hospital network. *Stat. Med.* 39 (8), 1125–1144.
- O'Malley, A. James, Bubolz, Thomas A., Skinner, Jonathan S., 2021. The Diffusion of Health Care Fraud: A Network Analysis. National Bureau of Economic Research. No. w28560.
- Opsahl, Tore, 2013. Triadic closure in two-mode networks: redefining the global and local clustering coefficients. *Soc. Network.* 35 (2), 159–167.
- Opsahl, Tore, Agneessens, Filip, Skvoretz, John, 2010. Node centrality in weighted networks: generalizing degree and shortest paths. *Soc. Network.* 32 (3), 245–251.
- Sah, Raaj K., 1991. Social osmosis and patterns of crime. *J. Polit. Econ.* 99 (6), 1272–1295.
- Shekhar, Shubhranshu, Leder-Luis, Jetson, Akoglu, Leman, 2022. Unsupervised Machine Learning for Explainable Medicare Fraud Detection. arXiv preprint arXiv: 2211.02927.
- Shi, Maggie, 2022. The costs and benefits of monitoring providers: evidence from Medicare audits. Available at SSRN: <https://ssrn.com/abstract=4063930>.
- Silverman, Elaine, Skinner, Jonathan, 2004. Medicare upcoding and hospital ownership. *J. Health Econ.* 23 (2), 369–389.
- Skinner, Jonathan, 2011. Causes and consequences of regional variations in health care. In: *Handbook of Health Economics*, 2. Elsevier, pp. 45–93.
- Skinner, Jonathan, Staiger, Douglas, 2015. Technology diffusion and productivity growth in health care. *Rev. Econ. Stat.* 97 (5), 951–964.
- van Capelleveen, Guido, Mannes Poel, Mueller, Roland, M., Thornton, Dallas, van Hillegersberg, Jos., 2016. Outlier detection in healthcare fraud: a case study in the Medicaid dental domain. *Int. J. Account. Inf. Syst.* 21, 18–31.
- Villalonga-Olives, Ester, Kawachi, Ichiro, 2017. The dark side of social capital: a systematic review of the negative health effects of social capital. *Soc. Sci. Med.* 194, 105–127.
- Welk, Paul, 2003. The corporate practice of medicine doctrine as a tool for regulating physician-owned physical therapy services. *J. Complex* 23, 231.
- Weng, Weifeng, Van Parys, Jessica, Lipner, Rebecca S., Skinner, Jonathan S., Sirovich, Brenda E., 2020. Association of regional practice environment intensity and the ability of internists to practice high-value care after residency. *JAMA Open* 3 (4) e202494–e202494.
- Zenou, Yves, 2003. The spatial aspects of crime. *J. Eur. Econ. Assoc.* 1 (2–3), 459–467.