RESEARCH ARTICLE

An empiric approach to identifying physician peer groups from claims data: An example from breast cancer care

Jeph Herrin PhD^{1,2} | Pamela R. Soulos MPH^{2,3} | Xiao Xu PhD^{2,4} | Cary P. Gross MD^{2,3} | Craig Evan Pollack MD, MHS⁵

¹Section of Cardiovascular Medicine, Yale University School of Medicine, New Haven, Connecticut

²Cancer Outcomes, Public Policy and Effectiveness Research (COPPER) Center, Yale University School of Medicine, New Haven, Connecticut

³Section of General Internal Medicine, Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut

⁴Department of Obstetrics, Gynecology and Reproductive Sciences, Yale School of Medicine, New Haven, Connecticut

⁵Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

Correspondence

Jeph Herrin, PhD, Section of Cardiovascular Medicine, Yale University School of Medicine, PO Box 2254, Charlottesville, VA 22902.

Email: jeph.herrin@yale.edu

Funding information

National Cancer Institute, Grant/Award Number: 1R01CA149045-01

Objective: To develop an empiric approach for evaluating the performance of physician peer groups based on patient-sharing in administrative claims data.

Data Sources: Surveillance, Epidemiology and End Results-Medicare linked dataset.

Study Design: Applying social network theory, we constructed physician peer groups for patients with breast cancer. Under different assumptions of key parameter values—minimum patient volume for physician inclusion and minimum number of patients shared between physicians for a connection—we compared agreement in group membership between split samples during 2004-2006 (T1) (reliability) and agreement in group membership between T1 and 2007-2009 (T2) (stability). We also compared the results with those derived from randomly generated groups and to hospital affiliation-based groups.

Principal Findings: The sample included 142 098 patients treated by 43 174 physicians in T1 and 136 680 patients treated by 51 515 physicians in T2. We identified parameter values that resulted in a median peer group reliability of 85.2 percent (Interquartile range (IQR) [0 percent, 96.2 percent]) and median stability of 73.7 percent (IQR [0 percent, 91.0 percent]). In contrast, stability of randomly assigned peer groups was 6.2 percent (IQR [0 percent, 21.0 percent]). Median overlap of empirical groups with hospital groups was 32.2 percent (IQR [12.1 percent, 59.2 percent]).

Conclusions: It is feasible to construct physician peer groups that are reliable, stable, and distinct from both randomly generated and hospital-based groups.

KEYWORDS

methods, patient-sharing, physician networks

1 | INTRODUCTION

Social network analysis has been increasingly used to examine patterns of health care delivery and outcomes.¹⁻⁷ In this context, pairs of physicians who provide care to the same patients have been shown to frequently know and refer to one another.⁸ The connections between pairs of physicians have been aggregated into larger connected groups of physicians ("peer groups") which may exhibit similar practice styles, potentially due to social influence, shared context, or other factors.^{9,10} There is growing evidence that physicians in peer groups deliver similar kinds of care, have similar rates of

complications following surgery, and follow similar patterns in adoption of new technologies.^{2,5,6,8}

Despite the increased interest in using physician peer groups in the application of network science to health services research, the identification of physician peer groups is inexact. A range of approaches have been used to construct the underlying physician patient-sharing networks, for example, using different assumptions about how many patients need to be shared and over what time period when defining a connection between a pair of physicians in a network. However, the impact of these essential assumptions on the properties of resulting peer groups has not been rigorously

evaluated. This is important because different assumptions used to identify patient-sharing between physicians could potentially have a profound impact on the reliability, stability, and inclusivity of the resulting peer groups and, in turn, their association with health outcomes.

Reliability of peer groups can be conceptualized as the degree to which physician assignment into specific groups is replicable. In a given time period, peer group assignment should be similar regardless of the specific sets of patients used to construct them. Moreover, peer groups should also be relatively stable over time, resulting in the assignment of the same physicians to the same groups in different time periods. Yet, little is known about how key decisions regarding the parameter values used in creating peer groups, including the number of patients that constitute a relationship between physicians and the number of patients a physician treats, might impact the reliability and stability of the resulting groups.

In addition to reliability, it is important to understand whether peer groups reflect entities that are meaningfully distinct from other ways of grouping physicians. For example, there are existing, simpler approaches to grouping physicians such as those assigning physicians to the hospitals where they practice or where their patients are admitted.¹¹ Peer groups constructed using shared patient algorithms should not simply recapitulate these hospital-based physician groups. If they are not distinct, they may contribute little additional information about patterns of care. Thus, even for reliable and stable peer groups, it would be important to understand to what extent they overlap with conventional physician groups such as those based on hospital affiliation.

To address these information gaps, we developed and tested an empiric approach for evaluating the reliability and stability of peer groups using a cohort of women with breast cancer and the physicians who treated them. We further evaluated stability of the resulting peer groups and compared them to hospital-based approaches to assigning physicians as well as randomly generated peer groups to determine the potential “added value” of network-based approaches. Though we used a specific cohort of patients, we mean to demonstrate an empiric approach to constructing and evaluating peer groups that could be applied to other patient populations in future studies.

2 | METHODS

2.1 | Overview

We used a cohort of Medicare beneficiaries with breast cancer to construct patient-sharing physician peer groups using a standard social network algorithm. First, during a single time period, we randomly divided patients into two cohorts and constructed peer groups separately for the same set of physicians using the two cohorts of patients. We systematically varied: (a) the minimum number of shared patients that constituted a link between a pair of physicians and (b) the minimum number of patients that a physician needed to treat to be included in the peer groups. For a range of values for both parameters, we assessed the reliability or the agreement of peer

group membership between the two sets of peer groups. Because higher thresholds for the minimum number of patients will tend to exclude more physicians, we examine the pattern of results for to identify parameters which balance the trade-off between good agreement (reliability) and maximum number of patients and physicians included (inclusivity). Second, using these parameters, we then constructed peer groups using all patients in two consecutive time periods and calculated agreement longitudinally, that is, comparing peer group membership stability over time. Third, for comparison, we constructed “control groups” by randomly assigning physicians to peer groups, and compared the stability of these randomly created groups to the stability of the empirically created peer groups. Finally, to determine whether these peer groups were distinct from conventional approaches of grouping physicians, we assessed the extent of overlap between physician peer groups and physician groups constructed using hospital referral patterns.

2.2 | Patient cohort

We used the Surveillance, Epidemiology and End Results (SEER)-Medicare dataset, which comprises data from regional cancer registries covering approximately 28 percent of the US population linked to Medicare claims data.¹² We also used contemporaneous Medicare claims for non-cancer patients residing in the SEER regions from the Medicare 5 percent random sample. We identified women diagnosed with breast cancer during 2004-2006 (“T1”) and 2007-2009 (“T2”), applying the following inclusion criteria: in situ or invasive stages I-III disease, first or only primary cancer diagnosis, histology consistent with epithelial origin, known date of diagnosis (“index date”), cancer not diagnosed on autopsy/death certificate, continuous enrollment in fee-for-service Medicare Parts A and B from 1 year before through 1 year after diagnosis or surgery if patient received surgery (or until death if patient died before 1 year had elapsed), had at least one claim billed to Medicare in the 1 year before through 1 year after diagnosis, and not diagnosed with a second non-breast cancer in the year following diagnosis/surgery. Women without cancer from the 5 percent sample were assigned a random “index date” during the time period which was used analogously to the date of diagnosis. All patients were assigned by zip code of residence to a single hospital referral region (HRR), as defined by the Dartmouth Atlas.¹³ Women without a valid ZIP code of residence were excluded.

2.3 | Physician cohort

For each woman in the sample, we identified any physician with a relevant specialty (radiologists, medical oncologists, radiation oncologists, surgeons, and primary care physicians including obstetricians/gynecologists) who billed a claim for treating her during the 3 months prior to through 9 months after their diagnosis/index date. Physicians were linked over patient claims and time using their National Provider Identifier (NPI). Because our timeframe spanned the switch in 2007 from the Unique Provider Identification Number (UPIN) to NPI, we used the UPIN-NPI crosswalk provided by NCI

for use with SEER-Medicare data to determine the NPI for claims in which this field was missing. Furthermore, NPIs can be used to identify both individual physicians and institutional entities. Because we were only interested in including individual physicians in our peer groups, we excluded any NPIs that linked to >1 UPIN in our Medicare claims, since these likely represented institutional NPIs. Physicians who treated patients from more than one HRR were themselves assigned to more than one HRR but treated as independently assigned (ie, as if they were two separate physicians) for the purpose of analysis.

2.4 | Peer groups

All peer groups were generated using the same algorithm, which was applied separately within each HRR. First, we restricted to those physicians who treated a minimum number of patients, denoted P . Then, we identified two included physicians as “connected” if they shared a minimum number of patients between them, denoted W (for weight of the connection). The resulting set of connections constituted a patient-sharing network. Figure 1 illustrates the full patient-sharing network for a single HRR; each circle represents a physician, and the lines between them (“edges”) indicate that they share at least one patient.

We then used the Girvan-Newman algorithm to disaggregate these networks into smaller subnetworks of physicians, which we term “peer groups”.^{14,15} This algorithm was selected because it has

been the most commonly used in health services research for physician patient-sharing networks.^{1,2,6} In this approach, physicians are placed in a single, mutually exclusive peer group. The intuition behind this approach is to progressively remove connections (“edges” or links between pairs of physicians) that are most likely to represent bridges between different groups while retaining the edges within a more tightly knit group. Formally, this is performed by calculating the “betweenness” score for each edge. The betweenness score of an edge is the number of shortest paths between all pairs of physicians that include the edge and incorporates the accompanying edge weight. A path represents the number of “steps” or physicians located in between a pair of physicians. In Figure 1, edges with greater betweenness are drawn with thicker lines. In each iteration, the edge with the highest betweenness score is removed and the betweenness score for each edge is recalculated. At each step, the network modularity is also calculated, which measures the proportion of edges that fall within the remaining clusters minus the proportion that would be in those clusters if the edges were distributed at random. The process is repeated until all edges have been removed; the step at which modularity was greatest is then retained as the solution.

2.5 | Selection of parameters to optimize reliability and inclusivity

To identify parameters that would optimize the trade-off between reliability and number of physicians retained (inclusivity), we split the T1 cohort into two subcohorts, T1a and T1b, by randomly assigning half of the patients for each physician to one subcohort or the other (note that physicians with only one patient in T1 were thus assigned to exclusively one subcohort or the other). Then, we constructed peer groups separately in each subcohort under a range of values for P and W . Specifically, we varied the minimum number of patients seen by each physician from $P = 1$ to 7 and varied the minimum number of shared patients for a connection from $W = 1$ to 4. Thus, for both T1a and T1b, we constructed a total of $7 \times 4 = 28$ sets of peer groups in each split sample.

In practical terms, increasing P (ie, minimum number of patients treated by a physician to be included) reduces the number of included physicians, but does not affect the direct connections between pairs of remaining physicians. For instance, if two physicians with at least seven patients are connected by sharing, they will still share patients when P increases and other physicians with fewer total patients are excluded. In contrast, increasing W (ie, minimum number of patients shared between two physicians for a connection) does not affect the number of physicians included, but reduces the number of physicians who are connected. For example, two physicians who share one patient would no longer be considered as sharing patients if W increases to 4. Thus, varying these two parameters across a range of values informs the patterns of trade-off between numbers of physicians included and the number of physicians connected.

For each set of solutions based on the 28 different permutations of (P, W) parameter values, we calculated reliability between the two

Full Network

FIGURE 1 Patient-sharing network for a single hospital referral region. Circles represent physicians; lines between two physicians indicate shared patients, with darker lines used for connections with greater betweenness. The five vertices on the right represent physicians who do not share patients with any other physicians in the same HRR, only with each other

TABLE 1 Number of patients and physicians included in the analysis

	T1	T2
	N	N
Patients		
Non-cancer patients	112 434	107 738
Cancer patients	29 664	28 942
Total	142 098	136 680
Physicians		
Medical oncologists	5681	6885
Primary care providers	48 111	51 513
Radiologists	29 188	31 626
Radiation oncologists	4614	4689
Surgeons	9128	10 859
Total	96 722	105 572

samples (T1a and T1b). To calculate reliability, we identified all physicians who treated patients in both T1a and T1b and had at least one other physician assigned to his/her peer group in T1a. For each of these physicians, we then calculated the percentage of other physicians in the physician's peer group in T1a who were also in their peer group in T1b. Thus, a score of 0 percent for a physician indicated that none of the physicians from his/her peer group in T1a were also in his/her peer group in T1b, whereas 100 percent indicated that all his/her T1a peer group members were also his/her T1b peer group members. For each combination of (P, W), we calculated the median and interquartile range of percent agreement for physicians in T1b. Physicians who were not placed in a peer group with any other physicians (termed "singletons") in T1a were assigned a score of 100 percent if they were also singletons in T1b, 0 percent otherwise. In addition to reporting the median and interquartile range (IQR) of reliability for each pair (P, W), we also graphed reliability against P for each value of W.

As mentioned above, the number of physicians who are eligible for analysis will vary according to values of P. Hence, for each combination of (P, W), we also calculated the number and percent of retained physicians (inclusivity). We then examined the relationship between median reliability and percent of retained physicians over the range of values for W and P to identify a pair of (P, W) values that optimally balanced the trade-off between reliability and the number of physicians retained. Conceptually, this optimal trade-off occurs when increasing either parameter results in only a trivial improvement in reliability.

2.6 | Stability of physician peer groups

Using the optimal values for P and W identified in the reliability analysis, we then created sets of peer groups using the same algorithm applied to all patients in T1 and T2, separately. We calculated stability of peer groups between different time periods using a metric similar to that used for reliability: For each physician in T1 who also treated patients in T2, we calculated the percentage of their T1

peers who were also T2 peers. As before, we summarized the median and IQR for the stability metric over all T1 physicians.

2.7 | Comparing stability to pseudo-peer groups

To further evaluate our stability results, we constructed "pseudo" peer groups by randomly assigning physicians to groups. That is, within each HRR, we randomly rearranged the physicians across peer groups, keeping the original size of each peer group the same. If, for instance, an HRR has only a single peer group comprising all physicians, the resulting pseudo group was the same as the original peer group; if, on the other hand, an HRR has many peer groups, the pseudo groups might have no agreement with the original groups. These were constructed in both T1 and T2; we then calculated their stability from T1 to T2 as above.

2.8 | Comparison of physician peer groups to hospital-based groups

As an additional comparison, we assigned all T1 physicians to hospitals according to where the plurality of their patients from our cohort had been admitted.¹¹ Then for each physician, we calculated the "overlap" between physicians assigned to the same hospital and to their peer group (based on the optimal (P, W) parameter values we identified). The overlap was defined as the number of physicians in both their assigned hospital AND their peer group divided by the number of physicians in either their assigned hospital OR their peer group. The overlap is 0 if there are no physicians in common, and 1 if they are all in common. We then summarized the overlap over all physicians in T1.

2.9 | Software

All analyses were performed using SAS 9.4 (SAS Institute Inc., Cary, NC), Stata 14.2 (StataCorp LP, College Station, TX), and R 3.2.1 (R Foundation for Statistical Computing, Vienna, Austria). The Girvan-Newman algorithm was implemented using *igraph* version 0.7.1.¹⁶

3 | RESULTS

3.1 | Sample characteristics

The study sample comprised 142 098 patients treated by 43 174 physicians in T1 (2004-2006) and 136 680 patients treated by 51 515 physicians in T2 (2007-2009), residing in 140 different HRRs. In our T1 split-sample analyses, there were 70 882 patients and 37 012 physicians in sample T1a and 71 216 patients treated by 37 335 physicians in sample T1b; of these, 29 133 physicians were in both samples. Characteristics of the physicians and their assigned patients are shown in Table 1.

3.2 | Selection of parameters to optimize reliability and inclusivity

The reliability of peer groups varied across different values of the two parameters P (the minimum number of patients treated by each

FIGURE 2 Reliability of group assignment between split samples for different values of minimum patients treated (P) and minimum number of shared patients (W) [Color figure can be viewed at wileyonlinelibrary.com]

Minimum # patients per physician (P)	Number of retained physicians	Percent of physicians not assigned to a peer group			
		Minimum # of shared patients (W)			
		1	2	3	4
1	37 012	16.4	25.0	51.6	65.6
2	28 074	16.4	15.4	36.9	55.4
3	22 410	16.4	11.8	26.4	44.3
4	18 436	18.4	9.7	19.5	35.0
5	15 425	17.5	9.1	15.2	27.2
6	12 913	17.9	8.8	12.0	20.7
7	10 983	18.3	8.7	9.6	15.8

TABLE 2 Number of retained physicians and percentage of physicians who are not assigned to a peer group ("singletons") for a range of values for minimum number of patients per physician (P) and number of shared patients (W)

physician) and W (the minimum number of shared patients for a connection). Figure 2 shows the relationship between reliability and P for each value of W. Holding the minimum number of shared patients constant while increasing the minimum number of treated patients generally improved the reliability (Figure 2). For example, using W = 3 to define a connection between physicians, reliability ranged from 66.7 percent in groups that required only a single patient per physician to be included in the network (P = 1) up to 87.0 percent for groups created using a minimum of seven patients per physician (P = 7; Figure 2). In contrast, holding the minimum number of patients treated fixed, while increasing the minimum used to define a connection between physicians, generally led to reduced reliability.

Increasing the minimum number of patients per physician resulted in fewer physicians being included in the cohort, ranging from 37 012 physicians when allowing for one patient per physician (P = 1) down to 10 983 when setting the volume threshold parameter at seven patients per physician (P = 7). As expected, increasing the minimum number of shared patients required to have a connection between two physicians did not impact the number of included physicians. However, a higher threshold for the number of shared patients did increase the number of "singleton" physicians (ie, those

with no other physicians in a peer group). This pattern of singletons is shown in Table 2; for W = 1, the rate was above 15 percent across all values of P. In contrast, for W = 2, the proportion of singletons declined with increasing P, from 25 percent to below 9 percent, as a larger proportion of physicians meet the threshold of at least two shared patients.

Based on these results, we identified P = 4 minimum patients per physician and W = 2 minimum shared patients as providing the optimal trade-off between reliability and inclusivity; increasing either value would not meaningfully increase reliability but would reduce the number of included physicians. Using these values, the median peer group reliability across physicians was 84.0 percent (IQR [0 percent, 95.2 percent]), with only 9.7 percent of physicians not being assigned to a peer group.

Figure 3 illustrates the peer groups created for the network in Figure 1 under the extreme and optimal values of P and W. In Figure 3A, the low minimum number of treated patients (P = 1) allows all physicians to be included, while the low threshold for a connection (W = 1) allows for the highest number of physicians to be connected to another. In contrast, Figure 3C for P = 7, W = 4 shows substantially fewer included physicians, as fewer treat at least seven patients; of these, several are not connected to any other physician, as the threshold of

FIGURE 3 Peer groups identified for the network in Figure 1 using three different sets of parameters. Each diagram A-C represents a distinct assignment of physicians to peer groups based on varying the parameters P (minimum number of patients) and W (minimum number of shared patients to create a link). Different shades represent different clusters of physicians identified by the model. Unshaded dots are physicians without the threshold number of patients P . The peer groups in (A) use all physicians with at least one patient ($P = 1$) and any connection of at least one shared patient ($W = 1$); in (C), only physicians with at least seven patient ($P = 7$) and connections of at least four shared patients ($W = 4$) are used. The peer groups in (B) correspond to the solution which the authors thought balanced the trade-offs best ($P = 4$, $W = 2$) [Color figure can be viewed at wileyonlinelibrary.com]

sharing $W = 4$ is more difficult to meet. Figure 3B shows the results of applying what we consider here to be the “optimal” parameters, that is, $P = 4$ and $W = 2$, in which a large proportion of physicians are included and connections between physicians retained.

3.3 | Stability of physician peer groups

Using $P = 4$ and $W = 2$, we constructed peer groups using all physicians in T1 and T2, separately. Median peer group stability from T1 to T2 was 70.8 percent (IQR [0 percent, 90.1 percent]). In contrast, the median stability over time for physicians randomly assigned to pseudo-peer groups was 5.7 percent (IQR [0 percent, 21.1 percent]).

3.4 | Comparison of physician peer groups to hospital assignment

We were able to assign 27 496 physicians from the T1 cohort (94.9 percent) to hospitals. For these, the median [IQR] overlap between hospital assignment and physician peer groups was 32.2 percent [12.1 percent, 59.2 percent].

4 | DISCUSSION

In this study, we have demonstrated that the property of physician peer groups varies with the parameters used to construct them, and propose an empiric approach to help address the problem which enables construction of peer groups that are reliable, stable, and distinct from hospital-based networks. These findings extend prior work in three important ways. First, we use an empiric split-sample approach to identify optimal parameter values for constructing

reliable peer groups; though this optimality is subjective, it does incorporate empiric results regarding how many physicians are retained and the percentage that are assigned to peer groups. Second, we found that, using these empirically derived parameters, the stability of peer groups over time was high. And third, we demonstrated that patient-sharing peer groups were distinct from groups based on hospital assignment, potentially providing unique and valuable information about how physicians relate and connect to each other beyond their hospital affiliation. These findings support both an approach to identifying peer groups and their potential to provide distinct information on health care delivery.

One key challenge for optimizing the health care system is understanding the dissemination of new treatments and technologies, especially those that are inefficient or unproven (or, conversely, understanding why proven treatments fail to disseminate quickly). A recent study of patients with breast cancer showed that the relationships between physicians are a key element in the spread of inefficient imaging practices by significantly influencing the adoption of perioperative use of magnetic resonance imaging (MRI) even after adjusting for hospital-based physician networks.⁶ Another study found that characteristics of patient-sharing networks explained rates of ambulatory care sensitive admissions.¹⁷ These findings suggest that physician networks may provide a context for interventions to help address overuse of low- or no-value medical services. Future research evaluating the potential effectiveness of such interventions will provide additional insights. The growing use of physician peer groups to study patterns of health care makes it even more critical to understand the best way to identify those peer groups. In our cohort, we found that reliability of constructed peer groups varied substantially depending on different parameter values used. In order to have reasonable reliability when constructing networks, physician

ties needed to be defined as having at least two patients in common and physicians needed to treat at least four patients. Though these parameters may differ for other cohorts, the method used to identify them can be generalized to other studies in which patient-sharing peer groups are constructed.

Our analysis extends current approaches used in the literature of physician peer groups, which primarily fall into three sometimes overlapping methods: (a) using an arbitrary minimum threshold for ties between physicians;¹⁸ (b) downweighting weaker ties;¹⁷ and (c) using a relative threshold, so that only ties in some top proportion are retained.¹ Notably, these approaches only concern the minimum number of shared patients; it is common to include all physicians, regardless of the number of patients seen. However, our split-sample analysis found that the number of patients seen (*P*) had a strong influence on the reliability of identified peer groups, independent of the minimum tie. This finding has important implications to other network analyses regardless of whether the split-sample approach is undertaken; restricting to higher volume clinicians will arguably always lead to more reliable patient-sharing networks. More importantly, these other approaches have selected the thresholds arbitrarily without evaluating reliability of the resulting peer groups; we propose a split-sample method that could be readily adopted in these other contexts, as well as other datasets and clinical focus, to better evaluate the impact of the thresholds used, not only on reliability but also on the number of physicians retained and the percentage who are singletons (not assigned to any group). Moreover, this split-sample approach can be used not only to optimize peer group construction using parameters we examined but also in other methods such as downweighting, or to optimize for different geographic areas or for peer group attributes other than reliability and stability.

We recognize that there is no convention for what constitutes the “optimal” trade-off between reliability and inclusivity, and thus, our approach does not provide a definitive solution to identifying the relevant parameters. However, our subsequent assessment of stability over time supports that our chosen parameter values performed well. Though the theoretical upper limit of stability is 100 percent, it is highly implausible that peer groups will be fully static, especially when looking at a national cohort. At the same time, our values of above 70 percent stability compare extremely well against the randomly generated groups with stability of 5.7 percent. Related to this is the question of validity; patient-sharing peer groups are a formal construct, and our results show they are distinct from hospital-based or randomly assigned networks. Hence, the choice of parameters may also be informed by how much the resulting peer groups explain patterns of care or other relevant measures. For instance, our prior work using a similar approach to select parameter values and construct peer groups found a strong association between peer group attributes and patient treatment, indicating at least some construct validity.⁶ However, in other contexts, the method used here to identify parameters may result in peer groups which do not have the same relationship with treatment or outcomes and thus may exhibit less construct validity. In particular, the current results are specific to the care of patients with breast cancer, a condition which typically

involves multiple specialties; in contrast, for patients with prostate cancer, for whom typically a single surgical oncologist provides care, the patterns of care and hence the stability of networks may vary substantially—we would expect the number of shared patients to be smaller in general. However, by including many patients without breast cancer in our sample, we hope to have provided results that are somewhat generalizable. While additional work in this area will produce more insights regarding what is acceptable or optimal, this is the first analysis we know of to evaluate cross-sectional reliability and longitudinal stability of patient-sharing peer groups and the values we found indicate what is possible and practical.

Our demonstrated evidence of good reliability of physician peer groups between cross-sectional split samples, high stability over time, and distinction from hospital-derived physician groups supports that patient-sharing peer groups are meaningful and distinct entities. This is consistent with our prior work, where we found that patient-sharing groups had independent influence on treatment patterns in addition to hospital groups.⁶ This finding has important implications for future research and policy discussion. For instance, research using hospital-based groups alone to examine patterns of care may be limited by the omission of peer group influences; at the least, attempts to either understand or influence physician patterns of care should consider the role played by patient-sharing as distinct from that played by hospital affiliation. Patient-sharing groups may provide an additional leverage point for affecting physician behavior and optimizing practices. By using actual patterns of patient-sharing, the approach may provide a way to map how changes in practice structure and affiliation—in particular, consolidation related changes such as increasing provider participation in accountable care organization and the shift of oncology practice to the hospital outpatient setting—alter the observed relationships, including referral relationships.⁸

There are several limitations to our work. We used a single algorithm, the Girvan-Newman, and did not test our approach using other algorithms.¹⁷ However, this is one of the most common and widely implemented algorithms, and the general approach to evaluating reliability can be easily adapted to other algorithms. Likewise, because we only included Medicare patients and physicians who treat those with breast cancer, our particular results may not be generalizable to other patient and physician samples. However, the overall approach presented in our study is broadly applicable, and further research replicating this method in other cohorts can provide additional evidence regarding the application of network methods. A conceptual limitation is that our approach still requires subjective judgement about what is an acceptable trade-off between inclusion and reliability; however, by explicitly quantifying these, it grounds this judgement in empirical results that would otherwise be lacking. A more practical limitation is that we stratified our peer group construction by HRR; this approach, as well as restricting to a single split sample versus using an approach based on selecting multiple samples, was used to reduce computing time. The choice of geographic boundaries may have an important effect on network structure.¹⁹ Finally, we assessed only two parameters, while others, such as a

relative threshold for edges, could also be varied. These parameters too can be investigated with similar assessment of reliability and stability as in our current approach.

In summary, we have proposed an approach for assessing the reliability and stability of physician patient-sharing groups and have used this approach to identify appropriate parameter values for an optimal trade-off between reliability and inclusivity. Applying the approach to a cohort of women with breast cancer, we have shown that patient-sharing peer groups can be constructed such that they are reliable cross-sectionally, stable over time, and distinct from hospital-based physician networks. These results support their use in assessing the role of physician peer groups in influencing health care delivery.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This project was wholly supported by a grant from the National Cancer Institute (5R01CA190017). Dr. Herrin has received additional support for unrelated research from the Centers for Medicare and Medicaid Services and Mayo Clinic. Ms. Soulos reports support from 21st Century Oncology for work which does not overlap with this work. Dr. Gross reports support for research endeavors distinct from this project, from 21st Century Oncology, Johnson & Johnson, and Pfizer. Dr. Xu has worked under contract with the Centers for Medicare & Medicaid Services to develop and maintain performance measures. Dr. Pollack reports stock ownership in Gilead Pharmaceuticals, which interest does not overlap with the current work.

ORCID

Jeph Herrin <https://orcid.org/0000-0002-3671-3622>

REFERENCES

1. Landon BE, Keating NL, Barnett ML, et al. Variation in patient-sharing networks of physicians across the United States. *JAMA*. 2012;308(3):265-273.
2. Landon BE, Keating NL, Onnela JP, Zaslavsky AM, Christakis NA, O'Malley AJ. Patient-sharing networks of physicians and health care utilization and spending among Medicare beneficiaries. *JAMA Intern Med*. 2018;178(1):66-73.
3. Lee BY, McGlone SM, Song Y, et al. Social network analysis of patient sharing among hospitals in Orange County, California. *Am J Public Health*. 2011;101(4):707-713.
4. Ong MS, Olson KL, Cami A, et al. Provider patient-sharing networks and multiple-provider prescribing of benzodiazepines.

J Gen Intern Med. 2016;31(2):164-171. Erratum in: *J Gen Intern Med*. 2016;31(5):588.

5. Pollack CE, Soulos PR, Gross CP. Physician's peer exposure and the adoption of a new cancer treatment modality. *Cancer*. 2015;121(16):2799-2807.
6. Pollack CE, Soulos PR, Herrin J, et al. The impact of social contagion on physician adoption of advanced imaging tests in breast cancer. *J Natl Cancer Inst*. 2017;109(8):1-8.
7. von Stillfried D, Ermakova T, Ng F, Czihal T. Patient-sharing networks: new approaches in the analysis and transformation of geographic variation in healthcare. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2017;60(12):1356-1371.
8. Barnett ML, Landon BE, O'Malley AJ, Keating NL, Christakis NA. Mapping physician networks with self-reported and administrative data. *Health Serv Res*. 2011;46(5):1592-1609.
9. Coleman J, Katz E, Menzel H. The diffusion of an innovation among physicians. *Sociometry*. 1957;20(4):253-270.
10. Iyengar R, Van den Bulte C, Valente TW. Opinion leadership and social contagion in new product diffusion. *Marketing Science*. 2011;30(2):195-212.
11. Bynum JP, Bernal-Delgado E, Gottlieb D, Fisher E. Assigning ambulatory patients and their physicians to hospitals: a method for obtaining population-based provider performance measurements. *Health Serv Res*. 2007;42(1 Pt 1):45-62.
12. SEER Medicare data. <https://seer.cancer.gov/about/overview.html>. Accessed December 5, 2017.
13. Dartmouth Atlas. <http://www.dartmouthatlas.org/>. Accessed December 5, 2017.
14. Girvan M, Newman MEJ. Community structure in social and biological networks. *Proc Natl Acad Sci USA*. 2002;99:7821-7826.
15. Newman MEJ. Modularity and community structure in networks. *Proc Natl Acad Sci USA*. 2006;103(23):8577-8582.
16. Csardi G, Nepusz T. The igraph software package for complex network research, InterJournal, Complex Systems 1695. 2006. <http://igraph.org>
17. Casalino LP, Pesko MF, Ryan AM, et al. Physician networks and ambulatory care-sensitive admissions. *Med Care*. 2015;53(6):534-541.
18. Pollack CE, Weissman G, Bekelman J, Liao K, Armstrong K. Physician social networks and variation in prostate cancer treatment in three cities. *Health Serv Res*. 2012;47(1 Pt 2):380-403.
19. Laumann EO, Marsden PV, Prensky D. The boundary specification problem in network analysis. In: Burt RS, Minor MJ, eds. *Applied Network Analysis*. London, UK: Sage Publications; 1983:18-34.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Herrin J, Soulos PR, Xu X, Gross CP, Pollack CE. An empiric approach to identifying physician peer groups from claims data: An example from breast cancer care. *Health Serv Res*. 2019;54:44-51.
<https://doi.org/10.1111/1475-6773.13095>