DR. JORDAN EVERSON (Orcid ID : 0000-0001-5355-6043)

Article type : Methods Article

Title: Comparing Methods of Grouping Hospitals.

Title, Full Name, Degree: Dr. Jordan Everson, PhD, MPP (Corresponding Author)

Affiliation: Department of Health Policy

Vanderbilt University School of Medicine, Nashville, Tennessee

Address: 2525 West End Avenue

Suite 1275

Nashville, TN 37203

Phone: (615)875-6207

Email: Jordan.Everson@Vanderbilt.edu

Title, Full Name, Degree: Dr. John M Hollingsworth, MD, MS

Affiliation: The Institute for Healthcare Policy and Innovation,

Ann Arbor, Michigan

Dow Division of Health Services Research,

Department of Urology,

University of Michigan Medical School, Ann Arbor

Address: 1500 E Medical Center Dr SPC 5330

Ann Arbor, MI 48109-5330

Phone: 734-936-7030

Email: kinks@med.umich.edu

Title, Full Name, Degree: Dr. Julia Adler-Milstein, PhD
Affiliation: Department of Medicine,
University of California, San Francisco
Address: 3333 California Street Suite 265
San Francisco, CA 94118

Phone: 415-476-9562
Email: Julia.Adler-Milstein@ucsf.edu

Acknowledgements:

Dr. Everson was supported by the Agency for Health Research and Quality (AHRQ) R36HS024454, University of Michigan, and Vanderbilt University Medical Center.

Dr. Adler-Milstein was supported by the University of Michigan and University of San Francisco.

Dr. Hollingsworth was supported by Agency for Healthcare Research and Quality 1R01HS024525 01A1 and 1R01HS024728 01 and the University of Michigan Medical Center.

Objective. To compare the performance of widely-used approaches for defining groups of hospitals and a new approach based on network analysis of shared patient volume.

Study Setting. Non-federal acute care hospitals in the United States.

Study Design. We assessed the measurement properties of four methods of grouping hospitals: hospital referral regions (HRRs), metropolitan statistical areas (MSAs), core-based statistical areas (CBSAs), and community detection algorithms (CDAs).

Data extraction methods. We combined data from the 2014 American Hospital Association Annual Survey, the Census Bureau, the *Dartmouth Atlas*, and Medicare data on inter-hospital patient travel patterns. We then evaluated the distinctiveness of each grouping, reliability over time, and generalizability across populations.

Principle findings. Hospital groups defined by CDAs were the most distinctive (Modularity=0.86 compared to 0.75 for HRRs and 0.83 for MSAs; 0.72 for CBSA), were reliable to alternative specifications, and had greater generalizability than HRRs, MSAs or CBSAs. CDAs had lower reliability over time than MSAs or CBSAs (Normalized Mutual Information between 2012 and 2014 CDAs=0.93).

Conclusions. CDA-defined hospital groups offer high validity, reliability to different specifications, and generalizability to many uses when compared to approaches in widespread use today. They may, therefore, offer a better choice for efforts seeking to analyze the behaviors and dynamics of groups of hospitals. Measures of modularity, shared information, inclusivity and shared behavior can be used to evaluate different approaches to grouping providers.

Key words. Healthcare markets, Hospitals, referral patterns, network analysis.

INTRODUCTION

Definitions of groups of healthcare providers are critical to a range of health policy and health services research efforts. Several methods exist to identify groups of interconnected hospitals; widely-used methods include geopolitical areas [such as Metropolitan Statistical Areas (MSAs) and Core-Based Statistical Areas (CBSAs)], and Hospital Referral Regions (HRRs).[1,2] Placing hospitals (or other providers) into inter-connected groups allows for assessment of whether and how hospitals collaborate in, communicate during or compete for the care of patients and which hospitals might work towards shared initiatives such as accountable care or health information exchange. Hospital groups defined by these methods have also been used to measure variations in treatment practices, provider organization, competition, workforce calculations, and other purposes.[3-8] However, no prior work has systematically assessed the dominant methods of defining healthcare groups against a common set of criteria to enable better-informed decisions about which to use.

Classical measurement theory highlights three key components of useful measurement on which these definitions can be compared: reliability, validity, and generalizability. **Reliable** measures are reproducible—reliable healthcare groups should have stable membership over time and under varied assumptions about how to group hospitals. The **validity** of a measure refers to how well it reflects an underlying 'true' value. Valid groups should clearly separate hospitals into closely linked subsets and minimize the connections that cross outside of the group,[9,10] and valid groups should update as market dynamics change. Finally, **generalizability** refers to the ability to extend measurement from the sample under study to a broader population.

Comparing the existing methods to define hospital groups serves to reveal important strengths and limitations of each. For instance, because HRRs have not been updated since their initial creation in 1993 and reflect referrals for specific types of surgery, they may be more useful for understanding long-term geographic trends in specialty care than MSAs, which change every ten years. However, HRRs may not be well suited to define hospital markets for those seeking to assess competitive dynamics because referral patterns for specialty care may not reflect broader inter-hospital competition. Similarly, MSAs may lack validity for some purposes for which they have been used because they are based on general travel patterns by employed individuals, not travel patterns for those seeking healthcare.

These limitations highlight opportunities for new definitions to complement existing ones. In particular, the development of community detection algorithms (CDAs) in the field of network analytics provides a promising alternative to define hospital groups. Unlike the widely-used definitions, CDAs leverage patterns of interactions between entities to define maximally distinct groups whose members are highly connected but share few connections outside the group, resulting in high potential validity. The value of this approach in healthcare has recently been demonstrated in studies that placed physicians and other organization into groups.[11-13] In these studies, physician groups have been defined by applying a CDA to networks based on physicians treating the same patients, reasoning that patient sharing may lead to similar physician practice patterns and reflect broad referral patterns. There is a similar need to study hospital practice patterns and referral patterns, but existing definitions of groups of hospitals, such as HRRs or MSAs, may not be well-suited for this need.[14]

Applying a CDA to available data on all Medicare Fee-for-Service patients shared between hospitals could, therefore, result in a conceptually appealing definition of hospital groups that complements the definitions in widespread use today.[15] With such a method, the reliability across varied measurement strategies and change over time can be easily assessed; validity may be high because available algorithms have been tested and validated in a wide variety of applications,[16,17] and the underlying data is based on the broadest feasible national group of patients. Finally, a CDA using a hierarchical method may provide a generalizable set of groups, since communities can be easily divided into smaller sub-groups or combined into larger groups, allowing for flexible application.

We therefore sought to compare three widely-used methods of defining hospital groups (HRRs, MSAs, and CBSAs), as well as a new CDA-based method, on the extent to which they produce reliable, valid and generalizable healthcare provider groups. Our results serve to inform a broad array of health services and health policy stakeholders about the differences between definitions of hospital groups and thereby support a more informed selection process. The measures we employ might also be adapted to compare procedures for grouping healthcare providers in other contexts.

OVERVIEW OF DEFINITIONS OF HOSPITAL GROUPS

For each method, we describe the underlying population or linkage on which groups are defined, and the rules used to divide hospitals into groups.

HRRs

HRRs define geographic areas based on patient travel patterns for specialty care. The *Dartmouth Atlas* group defined HRRs in two steps using 1993 Medicare Claims. They first identified the city in which each Medicare Fee-for-Service patients in each ZIP Code received hospital care, and then created Hospital Service Areas (HSA) by grouping ZIP codes by the city where the plurality of patients received hospital care. They then grouped HSAs into HRRs by identifying the city in which the plurality of Medicare patients from each HSA received hospital care *for major cardiovascular surgery and for neurosurgery*. Finally, some HSAs were reassigned to create geographically contiguous HRRs. HRRs have not been updated since 1993.

CBSAs & MSAs

CBSAs and MSAs define geopolitical areas and were created for general purposes, not purposes specific to healthcare. CBSAs are defined by the United States Census Bureau based on work-commuting travel patterns of employed populations. Following the decennial census, the Census Bureau identifies urban areas with a population of at least 10,000, and groups all counties that contain that urban area with counties in which at least 25% of the population either commutes to or from the core urban area for work. CBSAs that contain at least one urban area of at least 50,000 are considered MSAs while those containing one of more than 10,000, but fewer than 50,000 are considered Micropolitan statistical areas.[18] MSAs and CBSAs are redefined following decennial censuses and are updated periodically between censuses.

CDA Communities

The CDA method is based on inter-hospital travel patterns for all fee-for-service Medicare patients. The movement of patients between hospitals captures a variety of reasons why hospitals may be linked – e.g., through shared populations, referral patterns, unintentional inter-hospital travel patterns (e.g. readmissions), and transfers. As a result, grouping hospitals based on patterns of patient sharing likely facilitates the study of a range of phenomenon – e.g., patient outcomes, practice patterns, collaboration efforts, and competition. The selected CDA algorithm, the Walktrap algorithm, begins with ‘random walks’ through the network, in which each move by the walker is determined by the volume of patients shared between hospitals, and it then computes a (non-geographical) distance measure between hospitals based on the likelihood that the walker visits pairs of hospitals.[16,19] A final grouping is selected that maximizes the distinctiveness (i.e., the modularity, defined as the proportion of shared patients within the groups vs. between groups, relative to chance) of the groups.

DATA AND METHODS

Data

We identified all acute care, non-federal hospitals using the American Hospital Association (AHA) 2014 Annual Survey. We combined this data with several other sources. We merged this data with earlier versions (2012 and 2013) of the AHA survey, 2007 and 2013 delineations of MSAs, and 2015 Hospital Compare data.

We also merged this data with network data files released by the Centers for Medicare and Medicaid Services (CMS). The "Physician Shared Patient Patterns" files were released from 2009 to 2014 and were derived from Fee-for-Service Medicare claims housed in the Integrated Data Repository [20]. These files contain information on all healthcare providers appearing on Medicare claims, including hospitals and other institutional providers. We identified 4,602 of 4,638 non-federal acute care hospitals in the AHA database in the Shared Patients data. Hospitals appear on patients' claims when they are listed as the organizational or institutional National Provider Identifier (NPI). Each observation within these network data consists of three key variables: the two providers that share patients (i.e. provider partners) identified by their NPIs and the number of unique beneficiaries for whom both providers appeared on a Medicare claim within 30 days aggregated over the course of the year. Provider partners that shared fewer than 11 unique patients over the course of a year are excluded from these files. We checked the validity and reliability of these data (reported in the technical appendix).

Definitions of Groups

Updated definitions for CBSAs, MSAs, and HRRs were already present in the AHA data. To define hospital groups using the new CDA approach, we applied Pons and Latapy's 'Walktrap' community detection algorithm, implemented in igraph in R, to a network composed of hospitals connected by the volume of patients shared between hospitals.[19,21] This algorithm has been shown to perform well across a variety of networks, and it is fully hierarchical so that the resulting structure can be divided into subcomponents.[16,22] Hospitals are first combined into groups (or "communities") by combining individual hospitals that have the lowest 'distance' as defined by the algorithm. These initial groups are then combined into larger groups by distance. This results in 'close' hospitals belonging to the same groups, which are themselves distant from other groups. A final grouping is selected that maximizes the distinctiveness of the groups in the network using a measure known as modularity.[9] We then leveraged the hierarchical structure of this method to divide groups into smaller subgroups and then recombined any singleton hospital communities into their largest partner community. We created definitions with 266 groups (the modularity maximizing result, which is presented as a map of the continental United States in the technical appendix as Figure A1), as well as definitions with 308 and 863 groups to mimic HRRs and CBSAs. More detail on this approach is available in the technical appendix.

Comparison of Methods

We sought to compare each definitional method across key dimensions of reliability, validity, and generalizability. Where possible, we tested the performance of each method for each dimension empirically. Where this was not possible, we present a conceptual assessment. We based several evaluations on how well each method reflected patient travel patterns, which both reflect market dynamics and are likely to influence hospital behavior.

<u>Methodological Reliability</u>

We evaluated whether the groups resulting from each method were robust to changes in grouping rules or underlying populations. We further evaluated the reliability of the Walktrap algorithm by comparing the resulting community structure to several similar CDA approaches, and to an approach that eliminated shared patient ties between hospitals that are more than 60 miles away.

<u>Reliability Over Time & Responsiveness to Change</u>

For each method, we sought to examine the similarity of hospital groupings over time. This was not possible for HRRs because they have not been updated since they were defined in 1996 (using 1993 data). For MSAs and CBSAs, we compared the 2007 and 2013 delineation of MSAs and CBSAs to evaluate how hospital groupings changed over time.[23,24] For CDA communities, we applied the algorithm to each year of the network data from 2012 to 2014. We then examined the reliability of community identification by comparing membership in these four different sizes of groups in 2014, 2013 and 2012. To characterize the similarity of CDA and MSAs/CBSAs over time, we first identified the percentage of hospitals that were not included in the groupings in each year but were present in others. For the subset included in every year, we used the normalized mutual information (NMI), a measure of the amount of joint information contained in group partitions that ranges from a low of 0 to a maximum of 1. While we expected some movement in group definition based on changing relationships between hospitals, we would find evidence for reliable groups if the NMI is close to 1.

We also conceptually evaluated how likely it is that each method would be responsive to change in how patients travel between hospitals. Low responsiveness would indicate that broad change in group dynamics (such as new hospital ownership, expansion, or system membership) that would likely change patient flow, would not be captured by these methods. To the extent that this is true, it represents a limitation of the method's ability to capture important group dynamics, such that the definition may not validly reflect groups for some applications.

Validity: Distinctiveness

To evaluate each method's success at dividing hospitals into distinct groups with minimal links to other groups, we applied a metric known as modularity, which is commonly implemented in network analysis, to the shared patient network data. Modularity is defined as the proportion of patients that are shared within groups as opposed to between groups relative to what would be expected given the number of patients shared with hospitals in each group. Modularity varies from -1 to +1, with 0 representing no better or worse than random. High modularity scores demonstrate high validity in groupings, and have been frequently used to evaluate different methods in the networks literature. We measured the modularity of groups for our CDA solution at the three levels described above, and similarly evaluated the modularity of HRRs. For MSAs and CBSAs, we measured modularity for only the hospitals included in the statistical areas.

Generalizability: Inclusivity & Flexibility

In the context of grouping hospitals, we characterized the generalizability of the four methods by assessing the extent to which they were inclusive of broad populations and flexible to varied applications. Specifically, we addressed three questions: (1) how inclusive is the population of patients that the method uses to define groups by each method? (2) how inclusive is the population of hospitals grouped by each method? (3) To what extent does each grouping method offer flexibility in group sizes, such that the method can be generalized to the widest range of future analytic purposes?

Similarity

While each method has a different basis for defining hospital groups and is applied to different data in different years, it is possible that the resulting hospital groupings are sufficiently similar that method selection is not important. This would be true if, for example, each method results in grouping geographically proximate hospitals and defines similar geographic 'cut-points'. Therefore, as a natural extension of the comparison of hospitals, we sought to examine how similar or different the groups produced by each method were by comparing hospitals in 2014 AHA data grouped together by HRRs, MSAs, CBSAs, and CDAs. To do so, we used the NMI to compare across grouping methods. Whereas we previously used NMI to evaluate reliability over time, (by comparing NMI within method applied to different years), here we compare the NMI across methods. Low NMIs would indicate that the choice of grouping strategy might influence the result of analyses using these methods, while high NMIs would indicate that choice of grouping method was not as likely to be influential.

Validity Extension: Shared Behavior

Given that one key motivation to define groups of hospitals is to identify similarity in practice patterns,[25,26] it may be useful to those interested in practice patterns to assess the extent to which definitions identify groups of hospitals that behave similarly. This idea is supported by the study of 'mindlines' in healthcare, which focuses on how providers that communicate with one another influence each other's behavior.[27,28] We, therefore, sought to explore differences in the extent to which groups of hospitals under each method shared similar behaviors. We did so by randomly splitting each community, HRR and statistical area in half and testing the correlation between the mean scores of each half on five performance measures. We selected performance measures from the 2015 Hospital Compare data to represent three important types of performance: efficiency, outcomes, and process of care. The five measures are (1) Medicare Spending Per Beneficiary, (2) 30-day All Cause Readmission Rates, (3) Mammography Follow Up Rates, (4) MRI Lumbar Spine for Lower Back Pain, and (5) total process scores.

RESULTS

Methodological Reliability

<u>HRRs</u>

Because HRRs are defined by travel patterns for patient populations receiving specific types of care, their methodological reliability may be low: defined groups may differ if referral patterns for other types of care were considered.

<u>MSAs & CBSAs</u>

CBSA and MSA definitions may be sensitive to changes in the proportion of commuters used to define a county as part of the geographic area from 25% to some other value. For instance, using a cutoff of 20% would likely increase the size of several MSAs and the 25% rule used by the Census bureau may be arbitrary.

<u>CDA Communities</u>

Because the CDA method is based on a continuous measure of distance, rather than a selected cut point, and included all Medicare patients, rather than patients with specific diagnoses, the Walktrap method avoids weaknesses associated with the other methods. When we compared the Walktrap algorithm to several similar CDAs, we found high levels of agreement, indicating that specific algorithmic choice did not strongly alter defined groups (full details in the technical appendix). Similarly, excluding long-distance patient sharing relationships resulted in smaller communities with high levels of mutual information.

Reliability Over Time & Responsiveness to Change

HRRs

Because HRRs have not been redefined, their reliability over time cannot be determined and the definition is not responsive to changes in underlying patient travel patterns. Use of a plurality rule to group ZIP codes to HSAs and HSAs to HRRs may make group definitions sensitive to small changes that cross that threshold, but insensitive to large changes away from it. For instance, if patients in an HSA are treated in central cities of two different HRRs, with 90% of patients treated in HRR A and 10% of patients treated in HRR B, a large shift in patient flow—from 90%-10% to 51%-49% would not alter HRR definitions. In contrast, if the HSA was instead split between two HRRs 51%-49% initially, a very small switch to 49%-51% would change the HRR definition.

MSAs & CBSAs

MSAs and CBSAs were empirically stable over time. 2,609 of 4,602 hospitals are within 364 MSAs in the 2007 delineation of MSAs. Of the 2,591 hospitals grouped into an MSA in 2013, 32 (1.2%) were not in a MSA in 2007. Of the 3,430 hospitals grouped within a CBSA in 2007, 86 (2.5%) were not in a CBSA in 2010 and of the 3,374 hospitals grouped in a CBSA in 2013, 33 were not in a CBSA in 2007 (0.9%). Because they are not based on patient travel patterns and use a specific threshold to group hospitals, this approach is unlikely to be sensitive to change in healthcare travel patterns.

CDA Communities

The amount of "shared information" over time (i.e. the amount of information known about grouping in one year by knowing the grouping in another) was high when comparing groups in 2012, 2013 and 2014, with NMI >0.93 in all years and across the three different group sizes) indicating reasonably reliable group identification over time. By including all fee-for-service Medicare patients, this method captures changes in patient travel patterns over time for a broader group of patients not limited to specific conditions.

Validity: Distinctiveness

The modularity of 306 HRRs was 0.75. The modularity of MSAs for the 2,591 hospitals within them is 0.83; when this is expanded to include all CBSAs, this encompasses 3,344 hospitals and modularity is 0.72. The modularity of the Walktrap communities was 0.86 at 266 groups (the modularity-maximizing solution), 0.84 at 308 groups, 0.63 at 863 groups (Table 1, Row 6). These differences in modularity are generally larger than those observed in studies where modularity is used to evaluate the performance of different grouping algorithms.[29,30]

Generalizability: Inclusivity and Flexibility

<u>HRRs</u>

HRRs may only reflect referral patterns for the neurosurgical and cardiovascular surgery Medicare populations used to define them and therefore not generalize to broader patient populations. HRRs cover the entire United States and therefore capture all U.S. hospitals. They are defined through a 3-level hierarchy with ZIP codes nested within HSAs nested within HRRs, therefore offering some limited generalizability to analytic purposes that require varied size groups.

<u>MSAs and CBSAs</u>

MSAs and CBSAs reflect workers commuting travel patterns, which may not be generalize to some important populations in healthcare (e.g. retired Medicare populations). MSAs and CBSAs only cover areas within commuting range of urban areas, containing 56% and 75% of hospitals, respectively. They have limited flexibility in terms of size: they are hierarchical in that they are groups of counties, and allow users to focus on either only MSAs, all CBSAs, or combined statistical areas, which are groups of geographically adjacent CBSAs.

<u>CDA Communities</u>

CDA Communities as we have identified them are defined by all Medicare patient movement, rather than a specific patient population like HRRs. The focus on Medicare may limit generalizability to individuals with other insurers; however, this is also true of HRRs. CDA communities cover the entire United States and, therefore, can be applied to the full population of hospitals. They offer a high degree of flexibility, as the communities can be split anywhere along the hierarchy.

Similarity

In 2014, each method defined moderately similar groups, with normalized mutual information over 0.85 in all cases (Table 3). HRRs and both the modularity-maximizing CDA and similar number CDA solution shared 0.88 NMI. In metropolitan areas, HRRs, CDAs, and MSAs were reasonably similar, with NMI over 0.91; CBSAs also produced similar groups.

Validity Extension: Shared Behavior

The split-half correlations of each method are reported in Appendix Table A4, along with their relative ranks. Hospitals grouped together by the 266 and 863 CDA community approaches had the best median split-half correlation across all five measures. Hospitals grouped by HRRs had more similar performance than any of the census-based measures, but were less similar than CDA communities.

A summary comparing the three grouping methods across attributes is presented in Table 2. CDAs were preferable to other methods across seven of eight criteria. HRRs were equivalent to or preferable to MSAs and CBSAs on six of eight methods.

DISCUSSION

In this study, we compared three existing methods of defining groups of hospitals along with a new method using CDA that captured patterns of shared patients for all Fee-for-Service Medicare beneficiaries. Hospital groups defined using a CDA were preferable across seven of the eight dimensions evaluated. Despite the more than 20 years since their creation and the limited patient population used to define regions, HRRs performed reasonably well on the dimensions where they could be quantitatively evaluated. Both CDAs and HRRs appear preferable to census-based areas because they offer higher validity and include all hospitals.

We implemented a number of metrics to evaluate the performance of each approach to grouping hospitals. We believe that these metrics will be valuable for assessing other approaches to grouping providers, and for selecting approaches that are most useful for specific analytic and policy purposes. For instance, provider group definitions that maximize the distinctiveness of each group may be most useful for identifying providers that are accountable for the health of a population, since distinctive groups imply that populations of shared patients are well contained within each group. Grouping approaches that exhibit a high level of shared behavior may be best employed in studies seeking to understand regional practice variation because they best capture groups that behave similarly and separate groups whose practices vary greatly. Highly inclusive groups may be most useful when studying broad collaborative initiatives like the spread of accountable care organizations.

There are two clear instances of trade-offs between the metrics that we employ. One trade-off is between modularity maximization, which generally implies identifying few large groups, and identifying smaller, potentially more actionable groups. As large, high modularity communities were split into smaller groups, the modularity of the group definitions decreased, meaning that a greater proportion of patients travel between groups. Analysts will have to consider whether a larger number of small groups is more meaningful for their analytic purpose despite being less self-contained. The second tradeoff is between over-time reliability and responsiveness to change. Employing measures that are reliable over time may make many analytic tasks easier (for instance, panel analysis where it is useful for each provider to be nested within a single group) but may be insensitive to changes in market conditions that should lead to differing groups, like hospitals joining a multihospital system and changing their key referral partners. While no one grouping approach is likely to fulfill all needs, increased use of community detection methods, and tools to measure these groups, could lead to greater availability of validated off-the-shelf methods, providing researchers with better options for their specific analytic task.

CDAs performed well on tests of reliability and validity across the metrics that we employed. Because they are explicitly based on inter-hospital patient sharing they are a logical basis for measuring outcomes related to interactions between hospitals and the professionals that staff them. Perhaps the most important advantage offered by the CDA method is the ability to define valid groups at multiple levels. For many applications, HRRs and the modularity-maximizing community solutions are likely larger than the ideal choice. For instance, HRRs are sometimes used to define competition between hospitals; however, they are likely not well suited to this purpose because they cover areas that are larger than the average hospital's catchment area. When possible, use of custom measures of competition derived from claims data is a preferable solution to any of the methods defined here;[31] however, when access to claims is not available, small CDAs (made up of relatively few hospitals, on average) are likely to more closely identify hospitals that compete with each other by covering a smaller geographic region.

Despite these appealing properties, CDAs were not without limitations and the most serious limitation was that community definitions changed somewhat over time. It is not clear whether this reflects the ability to adapt to important changes in the underlying hospital network or low reliability over time. Regardless, changing group definitions could lead to analytic complexity over long time periods. In consequence, CDAs are likely the most useful option for short term or cross-sectional analysis because they offer the highest overall validity. Hospital to CDA crosswalks, defined at five levels of the hierarchy, are available as a technical appendix and at www.healthcareneighborhoods.net.

HRRs also performed well on most measures of validity and cohesion, though they lagged somewhat behind CDAs on most measures. However, because HRRs are large and not easily divided, they may be most useful when considering interactions focused on quaternary care hospitals or highly specialized care. Given their stable definition, they may also be useful for long panel data where stable definitions simplify analysis. An example of a study for which HRRs might be well suited is changes in the geographic variation in use of robotic surgery for neurosurgical care. However, HRRs are likely less useful when considering more local dynamics like referrals for simpler services, ED frequent fliers or similar dynamics.[32]

In contrast to CDAs and HRRs, we find little support for use of statistical areas for measuring issues related to hospital care. While CDAs and HRRs are defined based on patient travel between hospitals, MSAs and CBSAs are defined by where individuals live and commute and have been used to group hospitals based on this delineation. Therefore, CDAs and HRRs are likely most useful to measure issues related to hospital collaboration, like transfer protocols, health information exchanges, accountable care organizations, bundled payments, and readmission reduction programs. In comparison, MSAs are likely useful for defining broad populations to assess the needs or population health of an area and may be useful means to group hospitals when analysis occurs within this context (e.g., when assessing the adequacy of hospital beds for a population or the availability of some services).

Limitations. Key limitations to our CDA definitions stem from reliance on publicly available data derived from Fee-for-Service Medicare claims. This data excluded inter-hospital relationships that included fewer than 11 unique patients, potentially altering group definitions. Nevertheless, this approach should capture most inter-hospital relationships based on FFS Medicare patients because the average pair included in the data contained 310 shared patients. An additional limitation of the data source is that because the data lacks clinical information, our communities the reliability of these definitions for specific clinical groups or specialties cannot be readily ascertained. Beyond limitations of the data, many approaches to defining groups exist, and we have selected one and report its performance across a range of methods, as well as the reliability of these groups to other approaches. However, our approach is likely not the optimal solution for all cases; instead, we offer a new option for researchers to consider, and the best group definition is likely to depend on the question examined. Like other modularity-maximizing CDAs, our definition of communities may be subject to a resolution limit and selecting larger than optimal communities. For this reason, we have provided community definitions made up of more, smaller communities at lower level of the hierarchy that might provide more useful communities for specific analytic purposes. Finally, and more broadly, our study was not able to employ quantitative comparisons of each method for each dimension, and therefore had to rely on some qualitative (and potentially subjective) assessments.

CONCLUSION

Compared to widely-used methods to define groups of hospital, CDAs exhibit stronger conceptual and distinctive validity, and community membership is more closely related to hospitals' behavior than is other group membership. Nonetheless, HRRs performed reasonably well on several dimensions, and our findings provide support for their widespread use over the past two decades. Our metrics also demonstrate how researchers might compare methods of grouping providers in other contexts. Together, our results serve to inform researchers and other stakeholders in selecting a grouping method methodology that produces the most appropriate clusters of hospitals for their purposes and provides a new set of groups that may be useful when existing methods do not identify appropriate groups.

1. Wennberg JE. *The Dartmouth Atlas of Health Care in the United States (incl. Diskette).* American Hospital Association; 1996.
2. Douglas TJ, Ryman JA. Understanding competitive advantage in the general hospital industry: Evaluating strategic competencies. *Strategic management journal.* 2003;24(4):333-347.
3. Weinstein JN, Lurie JD, Olson P, Bronner KK, Fisher ES, Morgan MTS. United States trends and regional variations in lumbar spine surgery: 1992–2003. *Spine.* 2006;31(23):2707.
4. Carey K, Burgess JF, Young GJ. Hospital competition and financial performance: the effects of ambulatory surgery centers. *Health economics.* 2011;20(5):571-581.
5. Goodman DC, Fisher ES, Bubolz TA, Mohr JE, Poage JF, Wennberg JE. Benchmarking the US physician workforce: an alternative to needs-based or demand-based planning. *Jama.* 1996;276(22):1811-1817.
6. Neprash HT, Chernew ME, Hicks AL, Gibson T, McWilliams JM. Association of financial integration between physicians and hospitals with commercial health care prices. *JAMA internal medicine.* 2015;175(12):1932-1939.
7. Alexander JA, Lee S-YD, Griffith JR, Mick SS, Lin X, Banaszak-Holl J. Do market-level hospital and physician resources affect small area variation in hospital use? *Medical Care Research and Review.* 1999;56(1):94-117.
8. Song Y, Skinner J, Bynum J, Sutherland J, Wennberg JE, Fisher ES. Regional variations in diagnostic practices. *New England Journal of Medicine.* 2010;363(1):45-53.
9. Newman ME. Modularity and community structure in networks. *Proceedings of the national academy of sciences.* 2006;103(23):8577-8582.
10. Fortunato S. Community detection in graphs. *Physics reports.* 2010;486(3):75-174.
11. Casalino LP, Pesko MF, Ryan AM, et al. Physician Networks and Ambulatory Care-sensitive Admissions. *Medical care.* 2015;53(6):534-541.
12. Funk RJ, Owen-Smith J, Landon BE, Birkmeyer JD, Hollingsworth JM. Identifying Natural Alignments Between Ambulatory Surgery Centers and Local Health Systems: Building Broader Communities of Surgical Care. *Medical care.* 2017;55(2):e9-e15.
13. Landon BE, Keating NL, Onnela J-P, Zaslavsky AM, Christakis NA, O'Malley AJ. Patient-sharing networks of physicians and health care utilization and spending among Medicare beneficiaries. *JAMA internal medicine.* 2018;178(1):66-73.

14. Kilaru ASW, Douglas J.; Karp, David N.; Love, Jennifer; Kallan, Michael J.; Carr, Brendan G. Do hospital service areas and hospital referral regions define discrete health care populations? *Medical Care.* 2015;53(6):510-516.
15. Yujie Hu FW, and Imam M. Xierali. Automated Delineation of Hospital Service Areas and Hospital Referral Regions by Modularity Optimization. *Health services research.* 2018;53(1):236-255.
16. Danon L, Diaz-Guilera A, Duch J, Arenas A. Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment.* 2005;2005(09):P09008.
17. Harenberg S, Bello G, Gjeltema L, et al. Community detection in large-scale networks: a survey and empirical evaluation. *Wiley Interdisciplinary Reviews: Computational Statistics.* 2014;6(6):426-439.
18. Office of Management and Budget. 2010 Standards for Delineating Metropolitan and Micropolitan Statistical Areas.
19. Pons P, Latapy M. Computing communities in large networks using random walks. Paper presented at: International Symposium on Computer and Information Sciences2005.
20. Center for Medicare and Medicaid Services. https://questions.cms.gov/faq.php?faqId=7977.
21. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal, Complex Systems.* 2006;1695(5):1-9.
22. Orman GK, Labatut V, Cherifi H. Comparative evaluation of community detection algorithms: a topological approach. *Journal of Statistical Mechanics: Theory and Experiment.* 2012;2012(08):P08001.
23. United States Census Bureau. 2007 ZIP code to 2006 CBSA.
24. United States Census Bureau. 2010 ZIP Code Tabulation Area (ZCTA) Relationship File Layouts and Contents. https://www.census.gov/geo/maps-data/data/zcta_rel_layout.html.
25. Wennberg JE, Gittelsohn AM. Small area variations in health care delivery. 1973.
26. Birkmeyer JD, Sharp SM, Finlayson SR, Fisher ES, Wennberg JE. Variation profiles of common surgical procedures. *Surgery.* 1998;124(5):917-923.
27. Wieringa S, Greenhalgh T. 10 years of mindlines: a systematic review and commentary. *Implementation Science.* 2015;10(1):45.
28. Gabbay J, le May A. Evidence based guidelines or collectively constructed “mindlines?” Ethnographic study of knowledge management in primary care. *Bmj.* 2004;329(7473):1013.
29. Pujol JM, Béjar J, Delgado J. Clustering algorithm for determining community structure in large networks. *Physical Review E.* 2006;74(1):016107.

30. Wang R-S, Zhang S, Wang Y, Zhang X-S, Chen L. Clustering complex networks and biological networks by nonnegative matrix factorization with various similarity measures. *Neurocomputing.* 2008;72(1-3):134-141.

31. Colla C, Bynum J, Austin A, Skinner J. *Hospital competition, quality, and expenditures in the US medicare population.* National Bureau of Economic Research;2016.

32. Gresenz CR, Rogowski J, Escarce JJ. Updated Variable-Radius Measures of Hospital Competition. *Health Services Research.* 2004;39(2):417-430.

Table 1. Size and Distinctiveness of Grouping Methods.

	306 HRR	364 MSA	898 CBSA	266 CDA Communities	308 CDA Communities	863 CDA Communities
Average Size	28	28	22	31	26	8
Minimum	2	1	1	2	2	2
Maximum	85	116	116	93	93	37
% of ties within	51.8%	63.3%	48.9%	61.9%	57.8%	27.9%
% of patients within	75.0%	84.7%	72.4%	87.2%	85.3%	62.0%
Modularity	0.75	0.83*	0.72**	0.86	0.84	0.63

N=4,602 for HRR, and CDA communities, 2,591 for MSAs and 3,344 for CBSAs.

Ties refer to the number of instances in which two hospitals share at least 11 patients.

* MSA Modularity defined only for hospitals within MSAs. When HRR and CDA methods used to define hospital groups for this subset, the modularity is 0.78 for HRR and 0.89, 0.88, 0.69 for the 266, 308 and 863 communities.

** CBSA modularity defined only for hospitals within CBSAs. When HRR and CDA methods are used to define hospital groups for this subset, the modularity is 0.76, 0.87, 0.86, and 0.66.

Table 2. Summary of Hospital Grouping Methods Performance on Eight Measurement Properties

Measurement Property	Criteria	Description	CDA	HRR	MSA	CBSA
Reliability	**Method-based reliability**	Changes in methodology (e.g. cut-off points) should not arbitrarily change hospital community membership.	◕	▬	▬	▬
Reliability	**Reliability Over Time**	Changes over time should not arbitrarily change hospital community membership.	◔	●	◕	◕
Validity	**Responsiveness to Change**	Communities should change to reflect changes in patient movement or risk becoming misleading over time.	●	◔	◕	◕
Validity	**Defined by patient travel patterns**	Communities should be defined by patient travel patterns because they reflect referral patterns, and communication, competition for referrals and define coordination needs.	●	▬	▼	▼
Validity	**Highly Distinct**	Approaches in which hospitals within the community are highly connected, with few connections to outside hospitals, are more meaningful than approach that divide highly connected hospitals into separate communities.	●	◑	◕	◔
Generalizability	**Largest feasible population**	Because community definitions are used in a range of research and policy applications, they should be based on the broadest possible population to reflect wide range of referral relationships and coordination needs, and should include as many hospitals as possible nationally.	●	◕	◔	◑
Generalizability	**Adaptable Number of Groups**	Any delineation into a specific number of groups could be arbitrary; a hierarchical approach allows for division of communities into greater numbers.	●	◔	◔	◔
Validity Extension	**Members exhibit shared behaviors**	Hospitals, and providers practicing at them, that are members of the same community should exhibit similar behavior because they learn from one another and develop communities of practice.	●	◕	◑	◔

● Quantitatively verified to have high performance, and, where comparisons are possible, the best performer of included approaches.

◕ Quantitatively tested with moderately high performance and 2nd best performer when comparison possible.

◑ Quantitatively tested with moderately low performance and 3rd best performer when comparison possible.

▭ Conceptual reasons to question performance but not quantitatively testable.

▼ Strong conceptual reasons to doubt performance, but not quantitatively testable.

◔ Quantitatively tested with low performance.

Table 3. Similarity of CDA, HRR and Census-Based Definitions

	266 CDA Communities	308 CDA Communities	863 CDA Communities	MSA	CBSA
HRR	0.88	0.88	0.86	0.92	0.88
266 CDA Communities		0.99	0.89	0.93	0.89
308 CDA Communities			0.9	0.93	0.9
863 CDA Communities				0.91	0.91
MSA					1

Similarity measured by Normalized Mutual Information (NMI). For MSA and CBSA, similarity is only assessed for hospitals that reside within the statistical area.