This minitrack provides a venue for researchers who are advancing and applying the technology of data mining with the goal of improving healthcare practice. The intention is to use the unique forum provided by HICSS to allow expression of practical and theoretical, academic and industrial insight on this topic. Our mix of papers is intentionally eclectic, bounded by the desire to extract patterns from large data resources and use those to produce new knowledge for analysts and decision makers in the health sector. In some cases we are exploring new data mining techniques, however, that is not essential – at least as much potential for healthcare improvement lies with novel application of existing methods.

The first paper “Developing a Data-Driven Method for Estimating Provider Penetration and Abusive Billing Practices”, addresses the task of identifying providers that appear to be generating more claims than expected given patient and provider attributes. A provider’s penetration into a market is a key variable in monitoring for suspicious behavior, and a two-stage model uses clustering to identify the provider’s market area and HCFA data sources to model provider penetration.

The second paper, “Patterns Extraction for Monitoring Medical Practices,” investigates application of relational patterns to the monitoring of hospital discharge abstracts for quality of care. They demonstrate application of relational patterns to identify poor hospitalization practices.

The third paper “Precursory Steps to Mining HCFA Health Care Claims” discusses the work that must be performed prior to the actual data mining. These tasks include: customer discussions, data extraction and cleaning, transformation of the database, and auditing (basic statistics and visualization of the information) of the data.

The fourth paper, “A Global Optimization Approach to Cluster Analysis in Medical Diagnosis and Prognosis.” applies a technique based on convex and global optimization to Breast Cancer databases and achieves improvements in diagnosis and prognosis prediction over previously reported results.

The fifth paper analyzes data from the Health Insurance Commission (the Australian equivalent of HCFA). In “Data Mining of Administrative Claims Data for Pathology Services,” a number of new features are identified for use in predictive modeling. These features are summarized, visualized and used as inputs for clustering and outlier detection methods.

The sixth paper, presents “Descriptive Modeling in Healthcare Supporting a Facility Location Decision via GIS-Based Market Visualization.” This paper illustrates the decision support power of combining publicly available and system-specific data with a Geographic Information System (GIS) for locating and sizing a proposed Neonatal Intensive Care Unit (NICU) within a system’s network of rural hospitals.

Finally, the seventh paper, “Empirical Norms as a Lever for On-line Support of General Practice,” explores improvement of data entry efficiency and practice quality in primary care systems. The paper applies Bayesian models to a large primary care database to produce adaptive menus and dynamic decision support alerts.

This diverse selection of papers demonstrates the range of possibilities in data mining for healthcare quality, efficiency and practice support. It shows the range of application timing from point-of-care to financial review to facilities planning. It shows the range of methods from graphical display to novel query pattern and cluster analysis techniques. In all cases, the papers leave us with the realization that the healthcare systems can be significantly improved through the continued improvement and use of these techniques. Clearly there is scope for a return visit to Hawaii in 2002!
Developing a Data-Driven Method for Estimating Provider Penetration and Abusive Billing Practices

Hans R. Dutt, Michael Pepper, Joseph Brenner, John Stewart, Jonathan Smith and Mark Zezza

Abstract

This study develops the framework for an abusive-billing detection system that can potentially be used for examining claims of Medicare home health agencies, skilled nursing facilities, hospital outpatient facilities and hospice facilities. This system, which utilizes Medicare claims and other data, is comprised of two components: (1) a single-linkage clustering algorithm that defines market areas based upon patient/provider travel patterns and (2) a logistic regression model that estimates a provider’s expected billing share given the market it is defined to operate in. The system estimates expected shares and compares these shares to the actual share that each provider claims. In this manner, the system identifies outlier-billing providers who, in turn, are envisioned to be correlated with abusive-billing providers. While evaluation of the system requires thorough investigation of providers that are identified to be outliers by the system (which has not been under taken yet), there are preliminary indications that the model can identify providers that excessively bill Medicare, given patient and provider attributes. The system has been developed by HCFA and is called the Service Area and Provider Penetration System (SAPPS).

I. Introduction

In 1996, aggregate US health care spending surmounted one trillion dollars, of which, the Medicare program financed over 203 billion dollars. While private insurance spending grew at a rate of only 3.2 percent, Medicare spending grew at 8.1 percent (Levit et al. 1998). To many, this is evidence that abusive and fraudulent billing practices are rampant in the Medicare program. In addition, some estimates indicate as that as much as ten percent of US health care expenditures may be attributable to abusive and fraudulent billing practices, primarily by doctors and other providers (HCFA 1998).

This has lead Federal and state governments to initiate efforts to curb excessive billing practices in the Medicare and Medicaid programs. Included in these efforts have been the growing development and use of fraud and abuse detection systems such as EDS’s Texas Medicaid Fraud and Abuse Detection System (MFADS) and IBM’s Fraud and Abuse Management System (FAMS) which are used by some Medicare insurance contractors.

These systems are highly sophisticated systems that analyze vast amounts of claims data via techniques such as neural network algorithms that attempt to identify providers that exhibit anomalous activity for further investigation. These systems are typically quite expensive and require a substantial amount of effort and expertise from the user, who generally needs to continually input information for the system.

In this paper, we introduce an approach of examining claims records and identifying providers that appear to be generating more claims than expected given patient and provider attributes. The system has the advantage of being relatively low cost to construct and implement. Further, the system is grounded in economic theory and can provide estimates with statistical reliability.

A provider’s penetration into a market is a key variable to identify outlier behavior which, in turn, is correlated with abusive and fraudulent behavior. Unfortunately, estimation of expected provider penetration is not a simple task. First, the estimate requires that the market in which the provider competes be initially identified. Second, it requires that the number of claims or beneficiaries each provider draws within a market reflect factors that are likely to influence

---

Hans Dutt is a financial economist with the Securities and Exchange Commission. Michael Pepper is a policy analyst with the Department of Health and Human Services. Joseph Brenner, John Stewart, Mark Zezza and Jonathan Smith are with the Health Care Financing Administration. The ideas expressed in this paper are solely that of the authors and do not necessarily reflect those of the SEC, HCFA, HHS, ResDAC or their respective staffs.
market share (e.g., firm size and distance of patients to the facility).

Estimating a provider’s penetration was accomplished by a two-stage method. First, a method of identifying the provider’s market area was initially developed. More precisely, a clustering algorithm was built that utilizes Medicare claims data and examines patient-to-provider flows to create empirical market areas (EMAs) across the US. Subsequently, using HCFA data sources, a model of provider penetration was developed which accounted for fundamental market share determinants (i.e., patient-to-provider distance, firm size, etc.).

The four provider types under consideration are Medicare home health agencies (HHAs), skilled nursing facilities (SNFs), hospital outpatient facilities and hospices covered under Medicare Part A. A notable Part A institutional provider that was not considered is inpatient hospital services. The reason for this is that much academic work has focused on these types of claims. Further, Medicare Part B (physicians and supplier) claims are not considered. However, the work done in this study can readily extend to these claim types as well. Thus this research is useful in two respects. First, it extends market area and market share modeling to the national level of analysis. Second, it synthesizes the research into a fraud and abuse detection system in a manner that has not existed before.

II. Methodologies of System Components
A. Estimating Service Areas

1. Relevant Literature

There has been little academic research estimating service areas for home health agencies, skilled nursing, hospital outpatient and hospice facilities. Given this research gap, we were forced to draw inferences from existing academic work that has primarily been interested in physician and hospital market areas. Early market area estimation efforts were comprised of techniques involving an assertion that the provider’s service area is within some geographic boundary in which the provider existed. Providers within the same geographic unit are generally considered competitors. Common geographic boundaries used included fixed radius, standard metropolitan statistical areas (MSA), and county geographical areas.

Although these techniques were relatively simple to implement, significant criticisms were levied by many researchers, including Garnick et al. (1987) and Phibbs (1993). Garnick et al. states that some geographic units like counties may be too small to properly capture a hospital’s service area. Conversely, some geographic units, like MSAs, may be too large. As a result, these areas may define hospitals as competitors when they do not service the same population. Garnick et al. also argues that potential measures of hospital service areas neglect other factors which affect consumer demand such as income and level of health insurance coverage. Fixed radius measures are criticized because they require assumptions on patient and physician willingness to travel, hospital centrality, and ignores geographic boundaries.

Other researchers, including Drosness and Lubin (1966) and Carpenter and Pleassas (1985), have attempted to estimate hospital service areas by using birth and death records. These researchers reasoned that travel patterns of people who give birth or die in hospitals are reasonably representative of the hospital’s general population travel.

Recently, with the greater availability of discharge and claims data systems and the decreasing cost of processing large data sets, patient-origin analysis which examine the flow between where patients reside and where they receive services, have become a feasible method of directly estimating health service areas. Makuc et al. (1985) cites a conclusion of the Graduate Medical Education Advisory Committee that functional medical service areas be used as assessing the availability of physician services and that the physician market areas by specialty be determined empirically, based upon patient-origin data. Recent patient-origin studies have been conducted by Garnick et al. (1987), Morrisey, Sloan and Valvona (1988), and Makuc et al. (1991).
The only national patient-origin attempt to create market areas was conducted by Makuc et al. (1991), where routine care health service areas were created for the entire United States for the Centers for Disease Control. This study was distinct from others in that they used patient-origin data in conjunction with an agglomerative hierarchical clustering technique to cluster counties into market areas.

2. Data

The Health Care Financing Administration maintains National Claims History files from which Medicare beneficiary claims can be derived for various provider types including patient hospital, outpatient hospital, home health agency, skilled nursing facility, and hospice services. These files contain information on the beneficiary’s address (including county and zip code) and the providers from which they had received services, as well as other pertinent claim information. HCFA also maintains a system, referred to as OSCAR, which contains survey information obtained from unannounced surveys of providers and suppliers of Medicare/Medicaid services. This ensures compliance with Federal and/or State guidelines. From this system, attributes of providers and their addresses were compiled. Information on the provider’s county of residence was merged with the claims data. This data was aggregated to create measures of patient flow. Specifically, data was extracted to examine SNFs, HHAs, hospices, and hospital outpatient facilities for 1994.

3. Clustering Methodology

In general, clustering analysis is an exploratory data analysis technique that helps guide in searching large datasets for structures of natural groupings. The eventual groupings are determined by (1) the similarity measure and (2) the clustering method. The similarity measures are based upon patient flows. The clustering procedure is highly customized and would fall in the class of a hierarchical agglomerative and single-linkage method of clustering.

This process utilizes the general framework of Makuc et al. (1991) but forms the market area through an algorithm more akin to Garnick et al. (1989). The clustering approach we developed attempted to improve on Makuc’s foundation in two primary ways. First, Makuc’s algorithm utilized a clustering technique where the number of output market areas had to be predefined. The second drawback with Makuc’s approach was that the clustering procedure was so resource intensive that it necessitated the U.S. being clustered in segments and fused together by various assumptions with which the authors themselves expressed dissatisfaction. The method utilized in this work addresses these problems by requiring prior specification of only the level of inter-county reliance necessary to link any two counties, thus eliminating the necessity for pre-specification for the number of output market areas. In addition, the algorithm could process all U.S. counties simultaneously, eliminating the need to make assumptions to recombine regional segments.

Although clustering can be performed at various levels such as census tract, zip code or state, the county level has been chosen as the fundamental unit of analysis because it is the smallest geographic level on which national statistics are generally maintained (Makuc et al. 1991).

With respect to medical care, counties can be thought of in terms of production or consumption. Providers of a service produce stays or, if applicable, visits. Aggregating the production of providers in a county yields the county's total production. Total consumption of a county is the total utilization of stays/visits of residents of a particular county regardless of which county they have received services from.

Given any two counties, i and j, consumption flow can be measured in two ways: (1) county i residents traveling to county j (C(i,j)); and (2) county j residents traveling to county i (C(j,i)). Alternatively, in service types such as home health where health professionals visit patients consumption flows can be interpreted at (1) county i residents drawing visits from county j producers (C(i,j)); and (2) county j residents drawing visits from county i producers (C(j,i)). To determine the relative magnitude of a consumption flow of residents of county i to providers in county j (C(i,j)), it is necessary to examine it in terms of the total consumption of...
the county \( i \) \( (C(i)) \), and total production of county \( j \) \( (P(j)) \). More precisely, \( C(i,j)/C(i) \) measures the degree to which county \( i \)'s residents depend on county \( j \). Similarly, \( C(i,j)/P(j) \) measures the degree to which county \( j \)'s providers rely on county \( i \) residents.

\( C(i,j)/P(j) \) is a variant of the relevance index and \( C(i,j)/C(i) \) is a variant of the commitment index, which has been used extensively in service area research in studies such as Bay and Nestman (1980) and Carpenter and Plesses (1985). They are also the same type of cross county flow measures used in work such as Makuc et al (1991).

For any two counties \( i \) and \( j \), \( C(i,j)/C(i) \), \( C(j,i)/C(j) \), \( C(i,j)/P(j) \) and \( C(j,i)/P(i) \) are examined to determine if at least one measure exceeds some pre-specified critical reliance level. If the critical reliance level is exceeded, then the counties are linked. Then, starting with an arbitrary county, any other county having significant linkages with it are added to the cluster. These new counties are then examined for their linkages and new counties, if they exist, they are added to the cluster. This process continues until no more counties can be added and an arbitrary county is selected to start a new cluster.

The cluster formation depends upon level that the pre-specified minimum critical reliance threshold is set. The greater the reliance levels, the harder it will be for counties to connect and the smaller the cluster formations will be. Therefore, the proper reliance levels must be set by examining the cluster formations that result. We define four criteria to evaluate the appropriateness of the minimum critical reliance levels:

1. Reasonableness of cluster formations – market areas should generally be comprised of contiguous counties, and reflect migration patterns and travel patterns.
2. Market area self-containment – services should be sought and provided to a high degree within the geographic entity.
3. Reasonableness of counties per market area – since travel time is one of the most significant costs for Medicare patients, the number of counties should not be so large as to make travel within the empirical market area prohibitive.

Since there are tradeoffs between these criteria with respect to the critical reliance level, it is necessary to strike a balance between them. For each type of provider, clusters were formed by setting various critical reliance levels and evaluating the resulting clusters based upon the above criteria.

### B. Estimating Market Shares

#### 1. Relevant Literature

In addition to a research gap existing for service areas, there have been no published studies relating to estimation of market shares for home health agencies, skilled nursing, hospital outpatient and hospice facilities. Hence, we were again forced to draw inferences from existing academic work that has primarily been interested in physician and hospital market areas.

A number of studies have been published which concentrated on probabilities of hospital selection by patients. These include: (1) Folland (1983) who developed a multivariate spatial interaction model using intercity data to predict market shares using 1977 South Dakota data for general medical-surgical hospitals; (2) Cohen and Lee (1985) who related the probability of hospital selection to factors such as patient-to-hospital travel time, hospital attractiveness factors, physician characteristics and patient characteristics using 1980 Rhode Island data; (3) Erickson and Finkler (1985) who studied factors which affected hospital market share which were under the control of the hospital using a multiplicative competitive interaction model calibrated on 1979 Southeastern Pennsylvania data; and (4) Garnick, Lichtenberg, Phibbs, Luft and McPhee (1989) who compared linear estimation techniques (generalized least squares) with a conditional logit maximum likelihood estimator using 1983 California catheterization and coronary angiography data. Generally, these studies were concerned with estimating the probability a patient will select a particular hospital in a geographic entity (such as census tract).

The methodology followed by each of these studies was similar. Generally, the dependent variable was defined as the proportion of patients in a given area choosing a particular hospital. The explanatory variables included patient, physician, and hospital characteristics. Common
explanatory variables included the range of services offered, the availability of beds, location of provider relative to patients, the type of facility ownership, facility quality, number of affiliated physicians, urban setting, and presence of specialty units. The models were generally from the class of multinomial logit models.

2. Data and Methodology for Estimating Provider Penetration

Consistent with previous research, four explanatory variables were considered most likely to explain market share were the distance-to-provider, range of services offered, number of beds, and staff size. The distance-to-provider has been found to be negatively associated with provider selection and, by far, the variable that offers the greatest explanatory power. Patient to provider distance was obtained by purchasing 1995 location-zip code data from an outside vendor (Claritas). This data set contained the centroid longitude and latitude coordinates for each zip code in the United States. A program was written to calculate every provider-patient zip code combination by market area. This data was then merged onto the claims and utilization data so each beneficiary record contained the straight-line Euclidean mileage distance to the servicing provider.

The number of beds and size of staff variables have been found to be positively related to provider selection. This data was extracted through HCFA’s M204 OSCAR system and is essentially a snap shot of the provider file at the time the extract was made. The range of services variable was also found to be positively related to provider selection and was constructed by counting the number of revenue center codes which each facility reported in the claims data.

If they were appropriate for the provider type, they were included as explanatory variables and assessed for the contribution to explaining the variability in the market share variable. Four estimation techniques were considered to predict a provider’s market share: (1) ordinary least squares (Folland (1983)), (2) the multiplicative competitive interaction (MCI) model pioneered by Massao Nakanishi (1974), (3) a logit transformation model, and (4) a logistic regression.

All models regressed the proportion of a provider’s market area claims on appropriate combinations of distance, beds, staff size and range on services. Under this application, the R-squared statistic is of limited used in assessing the models. Therefore, we have chosen to evaluate the models by examination of the average absolute deviation of the predicted market share from the actual market share across providers.

III. Results

A. Finding Appropriate Empirical Market Areas

Clustering techniques require the user to specify a set of criteria in order to form appropriate clusters for a given application and, therefore, inherently require some subjectivity on the part of the user. In this application, we felt that two primary and competing criteria were of paramount importance. First, we wanted the degree to which care was sought and provided for within a market area to be as high as possible. This naturally would lead to the preference of larger market areas. On the other hand, we wanted the travel distance within the market area to be reasonable to reflect the fact that distance is an extremely important factor in patients’ choosing their health care provider.

Using these criteria, as well as subsequently checking that the market area formations made intuitive sense, sensitivity analyses of varying minimal critical reliance levels were used to select the following:

- For Medicare home health agencies, the minimum critical production and consumption reliances were chosen to be 30%.
- For Medicare hospital outpatient facilities, the minimum critical production and consumption reliances were chosen to be 20%.
- For Medicare hospices, the minimum critical production and consumption reliances were chosen to be 30%.
- For Medicare skilled nursing facilities, the minimum critical production and consumption reliances were chosen to be 20%.
In discussing the results, we use the following definitions:

(4) Isolated County Cluster - A county which had significant utilization and was defined by the clustering algorithm to be a market area in and of itself.

(5) Multi County Cluster - A group of counties, not necessarily physically connected, which the clustering algorithm identified to a market area.

1. Propensity of Self-Containment

Table 1 reports the self-containment measures at the selected minimum critical reliance levels.

For the established empirical market areas, the number of claims that remained within the cluster formation was high. On average, within market areas that were configured to contain more than one county, patients had from 82.1 (hospital outpatient) to 90.0 (hospice) percent of their claims serviced in the empirical market area. Likewise, within market areas that were configured to contain more than one county, provider’s obtained from 82.5 (skilled nursing) to 85.3 (hospice, home health agency) percent of Medicare claims from within the empirical market area. As expected, larger multi-county empirical market areas exhibited greater self-containment than did counties that were determined to be their own market area. However, the average self-containment of isolated counties was still quite high from both the consumption and production perspectives.

1. Size of Market Areas

Research has consistently found travel time to be an important cost a patient faces when choosing a provider. The fact that Medicare beneficiaries are generally elderly and their health care coverage is fairly comprehensive implies that travel time is likely to be a dominating cost. Hence market areas should be small enough to reasonable travel in, even at the expense of self-containment.

Table 2 reports size and related attributes of empirical market areas. It is recognized that sizes of counties vary significantly across the U.S. However, in general, it is felt that there is a significant relation between size of the empirical market area and travel time.

The empirical market areas that have been defined contain a significant number of market areas that are comprised of single counties. For the market areas that were defined by more than one county, the average number of counties per market area was under five across all provider types.

B. Finding the Appropriate Model

Once market areas are defined, the next step is to define a model to estimate expected provider market share. As noted previously, the variables that were considered important in determining provider selection were patient-to-patient distance, range of services offered, staff size and number of beds. Of course, all variables were not appropriate for all provider types.

The literature points to several potential models to estimate provider selection. These included: (1) include the ordinary least squares (OLS) such as used by Folland (1983), (2) logit transformation, multiplicative competitive interaction (MCI) used by Garnick et al (1989), and (3) the linear logistic regression model.

To examine which model was the most appropriate for this application, the four models were tested on 1996 Medicare skilled nursing facilities data. Results are summarized in Table 3.

For each model, within each market area, provider shares were estimated. Market share estimates were then aggregated to the provider level to yield expected market share and subtracted from the actual market share of the provider. The absolute value was then taken and the distribution of this variable examined to suggest the extent to which the model, on average, deviates from the actual market shares.

The ordinary least squares model is simple to employ and not greatly resource intensive. Empirically, it is accurate with a mean average absolute deviation of actual from expected market share of 1.6 percent. However, it possesses the theoretical problem of the estimates not being bounded to the zero to one probability space.

The logit transformation and MCI model have been suggested as methods to overcome this problem. The logit transformation model transformed the proportion of patients utilizing a
facility at a given distance to a logit (natural log of the quantity of the proportion divided by one minus the proportion) and a regression was subsequently employed. In the MCI model, proportions were transformed via a geometric transformation. In other words, the proportions of claims which providers served at various distances were divided by the clusters' geometric means and regressed on combinations of distance, beds, staff size, and range of services by service. While both models overcame the OLS problem, they appeared to be less accurate in practice.

In the final model under consideration, the proportion of claims of a cluster a provider served at a given distance was regressed on combinations of distance, beds, staff size, and range of services by service area utilizing a linear logistic regression. This approach computed estimates by the method of maximum likelihood. Although this model was not specifically cited in the literature, Garnick et al. (1989) demonstrated how maximum likelihood estimators were more stable under certain circumstances (many zero flows). Although more resource intensive, the model appeared to be slightly more accurate than OLS and its estimates were confined to the 0-1 probability space. Hence the logistic regression appears to be the most appropriate model of provider share.

Testing the Model:

We have developed a framework for a system that could be used to identify outlier-billing practices. The system developed from this framework, the Service Area and Provider Penetration System (SAPPS), has been created for HCFA and is currently in the early stage of testing. The system has been implemented and identified providers whose actual market penetration is greater than what the model would predict given the attributes of providers and patients. Ideally one would like to investigate the billing practices of providers that exhibit the greatest excess provider penetration in order to determine how well the system performs. Unfortunately, this has not been done as of this writing.

We have, however, gathered indirect evidence of whether this system has the potential to detect abusive providers. The Program Integrity Group in the Office of Financial Management at HCFA maintains a database, known as the Fraud and Investigations Database (FIDS). This database includes complaints on hospice, skilled nursing, and home health providers. For these provider types, we identified the providers in the FIDS system and examined their excess provider penetration estimates. In theory, if the model had randomly assigned excess penetration values, we would have expected that the average of these excess penetration estimates of FIDS providers to be zero. However, there are reasons to believe that, even if the model were statistically identifying abusive-billing providers appropriately, the mean of the excess penetration values would be near zero in this test. Further, many of the complaints may not be related to excess billing practices. To the degree that other types of complaints exist in this database, the test will be weakened. It should be noted however that it is difficult to isolate alleged abusive providers from other providers because of the structure of the database. Further, even if we could identify abusive-billing providers, the test will be weakened to the degree that the allegations were false. Table 4 presents the results of this analysis.

For all provider types, the investigated providers exhibited markedly more pronounced excess provider penetration than did the rest of the providers. Thus, despite the weakness of the test, there is indication that there is some positive association between providers being in the FIDS database and the model’s prediction of excess provider penetration.

IV Conclusion

This study makes several contributions from both an academic and practical standpoint. First, much of the academic literature has focused exclusively in the areas of hospitals and physicians in the areas of market share modeling and market areas. This study verifies that many of the previous findings in these areas are applicable to less well-studied but increasingly important areas of home health, hospice, skilled nursing and hospital outpatient. Most importantly, patient to provider distance has been found to be the most important variable in explaining provider selection.
Second, the work developed a clustering algorithm that can readily define market areas from claims data at a national level. The implications of this work go far beyond this application of a fraud and abuse detection system. For example, the same methodology, with slight adjustment can be used to define consumer market areas that can be used define Medicare HMO market areas or provider areas that can be subsequently used to examine provider competition in areas.

Finally, the paper defines a system that is less costly than those currently implemented and has quantifiable statistical reliability. While the capability of the model to detect providers with outlier billing and possibly abusive billing has not been fully evaluated to this point, we provide indirect evidence that suggests promise that this system can detect excessive billing to some degree. Given this, we recommend that HCFA proceed to the next phase of testing and investigate providers identified by the system to examine whether abusive practices are actually occurring.

References


Erickson, G.M., S.A. Finkler. “Determinants of Market Share for a Hospital’s Services.” *Medical Care*, 23.8 (August 1985)


Folland, S.T. “Predicting Hospital Market Shares.” *Inquiry* 20 (Spring 1983) :34-44


Health Care Financing Administration. 1998. Transcripts of the HCFA National Fraud, Fraud, and Abuse Conference


Table 1. Self-Containment of Market Areas at Selected Reliance Levels
Average Percentage of Claims Occurring Within Market Areas
1994 Medicare Fee-for-Service Claims Data

<table>
<thead>
<tr>
<th></th>
<th>Consumption</th>
<th>Production</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Multi-County Market Areas</td>
<td>Isolated County Market Areas</td>
</tr>
<tr>
<td>Home Health Agencies</td>
<td>84.9</td>
<td>75.1</td>
</tr>
<tr>
<td>Hospital Outpatient</td>
<td>82.1</td>
<td>70.0</td>
</tr>
<tr>
<td>Hospices</td>
<td>90.0</td>
<td>85.9</td>
</tr>
<tr>
<td>Skilled Nursing Facilities</td>
<td>82.7</td>
<td>72.8</td>
</tr>
</tbody>
</table>

Table 2. Size and Related Attributes Market Areas at Selected Reliance Levels
1994 Medicare Fee-for-Service Claims Data

<table>
<thead>
<tr>
<th></th>
<th>Average Number of Counties in Multi-County Market Areas</th>
<th>Number of Multi-County Market Areas</th>
<th>Number of Isolated County Market Areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home Health Agencies</td>
<td>4.65</td>
<td>453</td>
<td>985</td>
</tr>
<tr>
<td>Hospital Outpatient</td>
<td>4.79</td>
<td>447</td>
<td>1,055</td>
</tr>
<tr>
<td>Hospices</td>
<td>3.93</td>
<td>284</td>
<td>448</td>
</tr>
<tr>
<td>Skilled Nursing Facilities</td>
<td>3.61</td>
<td>427</td>
<td>944</td>
</tr>
</tbody>
</table>
Table 3. Self-Containment of Market Areas at Selected Reliance Levels
Average Percentage of Claims Occurring Within Market Areas
1996 Medicare Fee-for-Service Claims Data

<table>
<thead>
<tr>
<th>Potential Models</th>
<th>Ordinary Least Squares</th>
<th>Multiplicative Competitive Interaction Model</th>
<th>Logit Transformation</th>
<th>Logistic Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Average Absolute Deviation of Actual From Expected Market Share (MAAD)</td>
<td>1.6%</td>
<td>3.7%</td>
<td>4.0%</td>
<td>1.4%</td>
</tr>
<tr>
<td>Interquartile Range of MAAD</td>
<td>1.8%</td>
<td>3.5%</td>
<td>4%</td>
<td>1.6%</td>
</tr>
<tr>
<td>Competitor Provider Shares Sum to One?</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Computational Intensity: Probability Estimates Confined to 0-1 Probability Space?</td>
<td>Low</td>
<td>Moderate</td>
<td>Moderate</td>
<td>High</td>
</tr>
</tbody>
</table>

Table 4. Average Mean Excess Provider Penetration Estimates
Of Providers in FIDS Complaint Database – 1998 Data

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of Providers</th>
<th>Hospice Mean Excess</th>
<th>Home Health Agency Mean Excess</th>
<th>Skilled Nursing Facility Mean Excess</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>2,116</td>
<td>-2.21%</td>
<td>9,764</td>
<td>18.64%</td>
</tr>
<tr>
<td>Not Investigated</td>
<td>2,000</td>
<td>-2.32%</td>
<td>9,514</td>
<td>18.14%</td>
</tr>
<tr>
<td>Investigated</td>
<td>16</td>
<td>12.11%</td>
<td>241</td>
<td>38.81%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13,998</td>
<td>-0.74%</td>
<td>13,944</td>
<td>-0.82%</td>
</tr>
<tr>
<td></td>
<td>54</td>
<td>19.59%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>