Clustering Minnesota Physician Clinics

The Influence of Data Collection on Statistical Conclusions

Steven W. Nydick

May 12, 2011

Contents

1 Introduction **4**

2 Descriptive Summaries **6**

3 Cluster Analysis **14**

4 Conclusion **23**

Bibliography **24**

List of Figures

1 Boxplots of mean observed rates by region 1 8
2 Boxplots of mean observed rates by region 2 9
3 Boxplots of mean observed rates by region 3 10
4 Four linear discriminant analyses using different datasets and observations 13
5 Coordinate projections of fourteen model-based clusters 16
6 Dendrograms of hierarchical clusters 17
7 Boxplots of mean observed rates by hierarchical cluster 1 19
8 Boxplots of mean observed rates by hierarchical cluster 2 20
9 Boxplots of mean observed rates by hierarchical cluster 3 21
10 Coordinate projections of four model-based clusters 22

List of Tables

1 Means and standard deviations of observed rates 7
2 Means of the observed rates in each region 11
3 Means of the observed rates in each region for complete data 11
4 Means and standard deviations in clusters based on diabetes and vascular 23
5 Means of the observed rates in each medical group 25

Abstract

Minnesota recently started requiring clinics and hospitals to report health outcomes, including the rates of performance for certain procedures. However, when reporting the performance of Minnesota physician clinics, all but two of the procedures were aggregated among clinics within a particular medical group. Presumably, aggregating within medical groups for less common procedures allows for comprehensive reporting without violating privacy concerns. Yet, the Minnesota Department of Health (MDH) only indicated the limitations of interpreting individual rates in appendices. Thus, most consumers will assume that all of the reported rates and confidence intervals apply to the individual clinics and not only to their associated medical group. The following report assesses whether the conclusions drawn from the aggregated dataset is similar to those drawn without assuming that every clinic within a given medical group is identical. I first grouped clinics together based on rates of performance over the entire set of procedures, and I compared the resulting cluster structure to that based on rates of performance for the two procedures allowed to vary among clinics within a given medical group. The two sets of analyses led to different interpretations about the structure of Minnesota clinics. For example, using the complete dataset, there appear to be striking regional difference in clinic performance, which disappears after removing the redundant rates. I ultimately recommend that the future reporting of clinic performance either take into consideration the variability of clinics within a given medical group (by providing more appropriate uncertainty estimates) or better stress and identify the limitations of the available data.

1 Introduction

In 2008, Minnesota developed laws to improve the quality of healthcare across the state and to "create a uniform approach to quality measurement in order to enhance market transparency" [5]. Statute 62U.02 includes two subdivisions that attempt to standardize the assessment and transparency of medical practices. Subdivision 1 emphasizes "uniform definitions, measures, and forms of submission" while making sure to "incorporate measures for primary care, including preventative services" [6], whereas subdivision 3 impels clinics and hospitals to submit patient outcomes so that the commissioner of health can "issue annual reports on provider quality" [6]. That is to say, Minnesota now requires hospitals and clinics to report outcomes based on standardized assessment measures. These laws were developed in the context of a healthcare reform bill passed in 2008 to improve health, patient experience and affordability of health care in the state of Minnesota [10].

Obligated to assess clinic practices, the Minnesota Department of Health (MDH) developed a set of reporting procedures and hired a team of data collectors (led by Minnesota Community Measurement) to carry out the assessments. The first set of administrative rules requires clinics to register each year, submit data, complete a technology survey, and agree to validation measures. The procedures are put in place to reduce the burden on the data collectors while keeping the clinics and hospitals honest as to their practices. Data for the 2010 report are accessible on the following webpage:

http://www.health.state.mn.us/healthreform/measurement/report/index.html

The Minnesota Department of Health (MDH) partitioned the complete report into two sections, measures assessing the performance of hospitals and measures assessing the performance of physician clinics. Each of the major reports was divided into four sub-reports targeting hospitals or clinics in particular regions of Minnesota. The regions were designated as follows: (1) Northwest Minnesota (NW); (2) Northeast Minnesota (NE); (3) the Twin Cities Metro Area (TC); and (4) Southern Minnesota (S). Within each of the sub-regions, the MDH attempted to reach a wide audience, including health care providers and general consumers. Reporting styles were chosen specifically to reach a consumer base. For instance, at the outset of a report, the MDH explained each of the measures, why they believed them to be important, and the range of values across all of the clinics or hospitals. Moreover, they indicated how consumers should judge performance. As an example, explaining the colorectal cancer measure, they described the population (all patients ages 51 – 80), the criterion (received one of the listed screening procedures), the importance of the criterion (screening is effective at catching colon cancer early), the range of performance (43% to 94%) and what consumers should look for in performance (higher is better).

In the clinic reports, the MDH proceeded to list clinics alphabetized within cities, which were alphabetized within region. For a given clinic, the MDH divided the assessments into three major types and displayed the risk adjusted rate[1] for each assessment. The three major types of clinic procedures are as follows:

1. Chronic illnesses: Adults with diabetes, adults with vascular disease, adults with high blood pressure, and children/adults with asthma.

2. Acute illnesses: Children with a cold, children with a sore throat, and adults with bronchitis.

3. Prevention: Breast cancer, cervical cancer, colorectal cancer, general cancer, chlamydia, and childhood immunizations.

As in the colorectal cancer example, a rate for a given procedure indicates the percentage of times that the clinic took the correct action according to MDH protocols compared with either a sample of relevant patients or the entire population of relevant patients. Thus, a *higher* rate indicates *better* performance, and the Minnesota Department of Health lists this statement on *almost* every page of the clinic reports. Two of the measures, optimal diabetes care and optimal vascular care, were reported directly from clinics. The remaining measures were reported based on medical group and linked to the appropriate clinics. The adjusted rates for all clinics and medical groups with relevant specialties were reported as long as the sample size relating to a given procedure exceeded 30 patients. Additional information, such as the method of data collection, sample size, payment method, observed rates, the source of the data, and confidence intervals relating to both the observed and adjusted rates were relegated to appendices.

Even though the method of presentation is appropriate for consumers with little knowledge of statistics, it is nearly impossible to draw broad conclusions or spot general trends solely from pages of individual rates. Determining the overall performance of clinics and hospitals across the state would help consumers make decisions in where to live or compare the performance in their region with the performance in other regions around the state. Moreover, by analyzing the relationship between covariates, such as payment method, and rates of performance, we might be able to broadly classify clinics and hospitals into groups and concisely describe the content of those groups. That is, finding a more cohesive organizational structure could provide consumers with clarity of thought and policy makers with information on where to target interventions.

The following report is intended to provide some analyses missing from the Minnesota Department of Health website. Taking the performance rates listed by the Minnesota Department of Health, I looked at general trends across all of the clinics and attempted to group clinics into

[1] They intended to adjust the rate for hospitals with patients more prone to low performance measures, but they used payment plan as a proxy for low performance. For example, if Clinic A has more commercial patients than Clinic B while Clinic B has more Medicare patients than Clinic A, the rates of Clinic B are adjusted in the direction of better performance to account for the population more susceptible to bad outcomes [5]. However, many of the adjusted rates are risk-adjusted mortality rates, which hospital quality researchers have argued as misleading and ineffective [7, 12].

clusters based on rates of performance. I chose to examine the clinic data rather than the hospital data for a simple reason: The percentage of missingness was 63.92% for the hospital data and only 19.28% for the clinic data. With fewer missing data, the clinics are more amenable to standard statistical analyses. Using the clinic data set, I tried several methods of imputing missing values, including multiple imputation [11], random forest imputation [4], and imputation based off of an EM algorithm [9]. Each of the imputation algorithms resulted in a different conclusion, and I ultimately decided instead to delete clinics that were not fully observed. A fundamental problem with deleting partially observed rows is that by virtue of containing missing data, those clinics might be *fundamentally different* than those fully observed. The reader should keep in mind that the following analyses only applies to a subset of the complete data (see Schafer & Graham, 2002 for a nice overview of both traditional and modern missing data methods [8]). Moreover, because they did not clearly indicate that most procedures were aggregated over medical group, and because the clinics were listed alphabetically within city and not by medical group, I did not perceive the data limitations until nearly completing the report. Thus, the following analyses contrasts conclusions based on medical group specific rates and those reported individually among all clinics. I conclude with a recommendation for the future reporting of clinic performance.

2 Descriptive Summaries

Before simplifying the structure of the clinic data, it is worthwhile to examine the original set of variables. I decided to analyze the observed rates rather than the adjusted rates because both sets of variables gave nearly identical results, and the relationship between payment plan, sample size, and rate might be more interesting than originally presumed to calculate the adjusted rates. Table 1 displays the means and standard deviations of the observed rates across all clinics with complete observations. Four of the five procedures with the highest mean observed rate are directed primarily at children, including asthma, the common cold, sore throats, and childhood immunizations. Furthermore, the lowest mean observed rates tend toward treatment of chronic illnesses directed at adults, including bronchitis, diabetes, and vascular issues. Two of the three lowest mean observed rates, diabetes and vascular, were the only two conditions individually reported by clinics and not aggregated over medical group. Figures 1 – 3 display boxplots of observed rates for clinic performance on all of the conditions broken down by region. Clinics within the twin cities metro area (TC) consistently performed the highest on all of the observed rates, whereas clinics within southern Minnesota (S) tended to perform relatively low. Table 2 displays similar information to Figures 1 – 3, comparing the mean observed rates across all of the regions on all of the conditions. Notice that the mean observed rates of the twin cities metro area are higher on *almost every* condition than the mean observed rates of the other regions. Yet, there are two caveats in interpreting the rates. First, the rates were only calculated on a subset

Table 1: The mean and standard deviation of observed rates of the matching MDH recommended procedure for acute and chronic treatments and preventative measures across Minnesota clinics rounded to two decimal places.

Measure	Mean Observed Rate	SD Observed Rate
Asthma	.91	.03
Common Cold	.88	.08
Sore Throat	.86	.10
Breast Cancer	.83	.04
Immunization	.82	.11
Cervical Cancer	.79	.04
Colorectal Cancer	.72	.13
Blood Pressure	.69	.11
General Cancer	.52	.16
Chlamydia	.50	.11
Vascular	.36	.12
Diabetes	.29	.12
Bronchitis	.17	.07

of the data containing fully complete observations. As a point of comparison, the mean observed rates using all available observations are displayed in Table 3. In almost every case, the mean rate drops when taking into consideration all of the observations, but the *relative* rates across all of the regions do not much change. Second, only for diabetes and vascular do clinics report individual rates. All of the other procedures are aggregated across medical group. An obvious consequence of aggregation is that many of the region specific boxplots are a straight line at the median with a few outliers. Table 3 also includes the standard deviations of the observed rates across all of the available data. All but one of the standard deviations increase from the standard deviations displayed in Table 1 using only the completely observed clinics. The clinics missing observations are less likely to be in a large medical group than those fully observed, and part of the increase in variability captures the additional, inter-clinic heterogeneity.

Because the general trend of Tables 2 and 3 suggests that similar regional differences exist across all of the procedures, one might wonder whether we can capture those differences by combining information from all of the observed rates. Moreover, because only two of the observed rates varied within medical group, a skeptical reader would inquire as to whether the general regional differences held to the same degree when looking only at those variables. To assess whether there were separable regional differences, I performed four linear discriminant analyses (LDA), and the first two dimensions of each are displayed in Figure 4. The results are simple and enlightening. First, the upper left plot includes all of the observed rates for the 206 clinics that were completely observed. As suggested in Table 2, TC tends to have higher rates across the board relative to S, and those differences are captured in the upper left plot. The first linear dimension is essentially

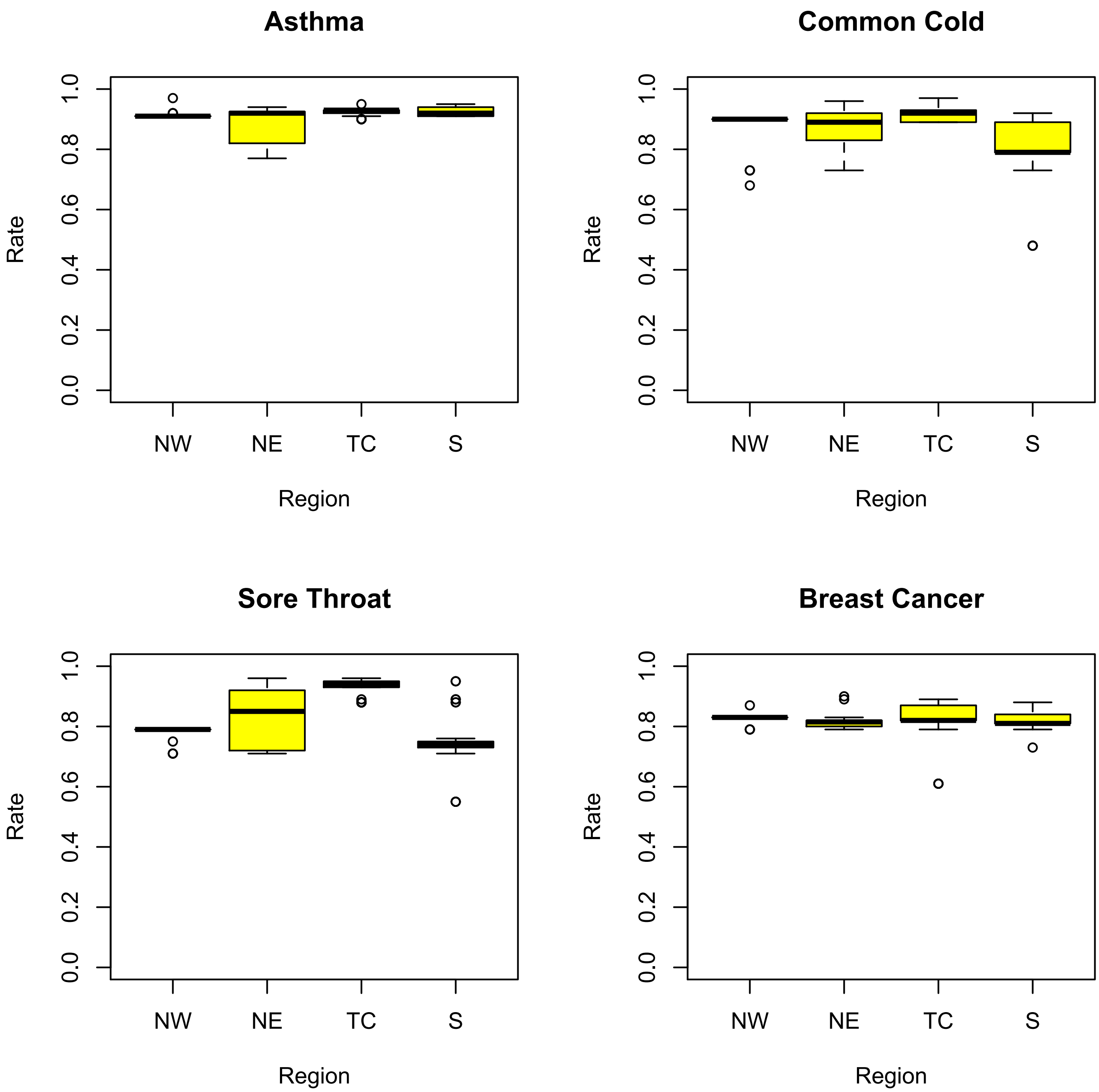

Figure 1: Boxplots of the four procedures with the highest mean observed rates, broken down by region.

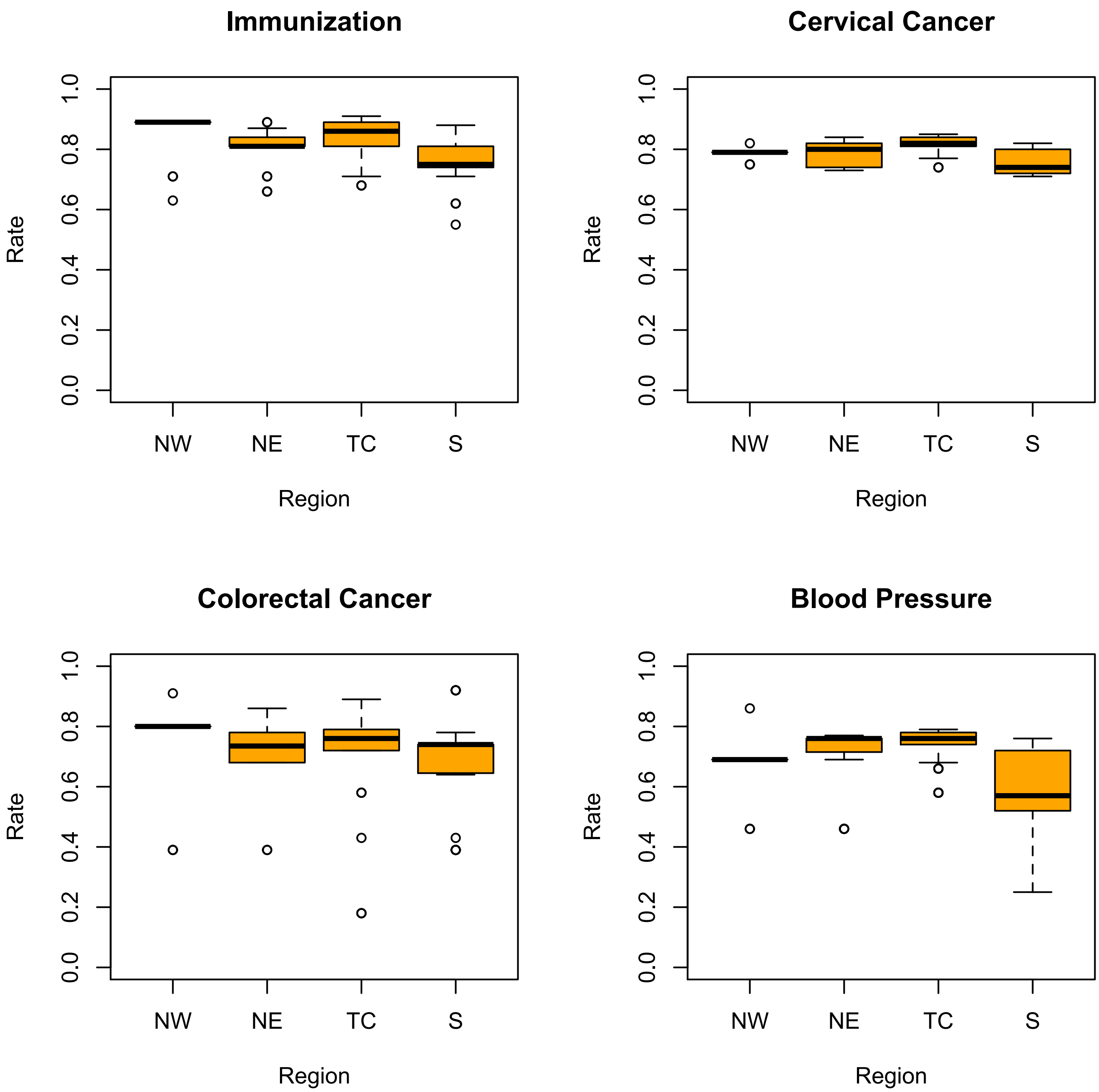

Figure 2: Boxplots of the four procedures with the middle mean observed rates, broken down by region.

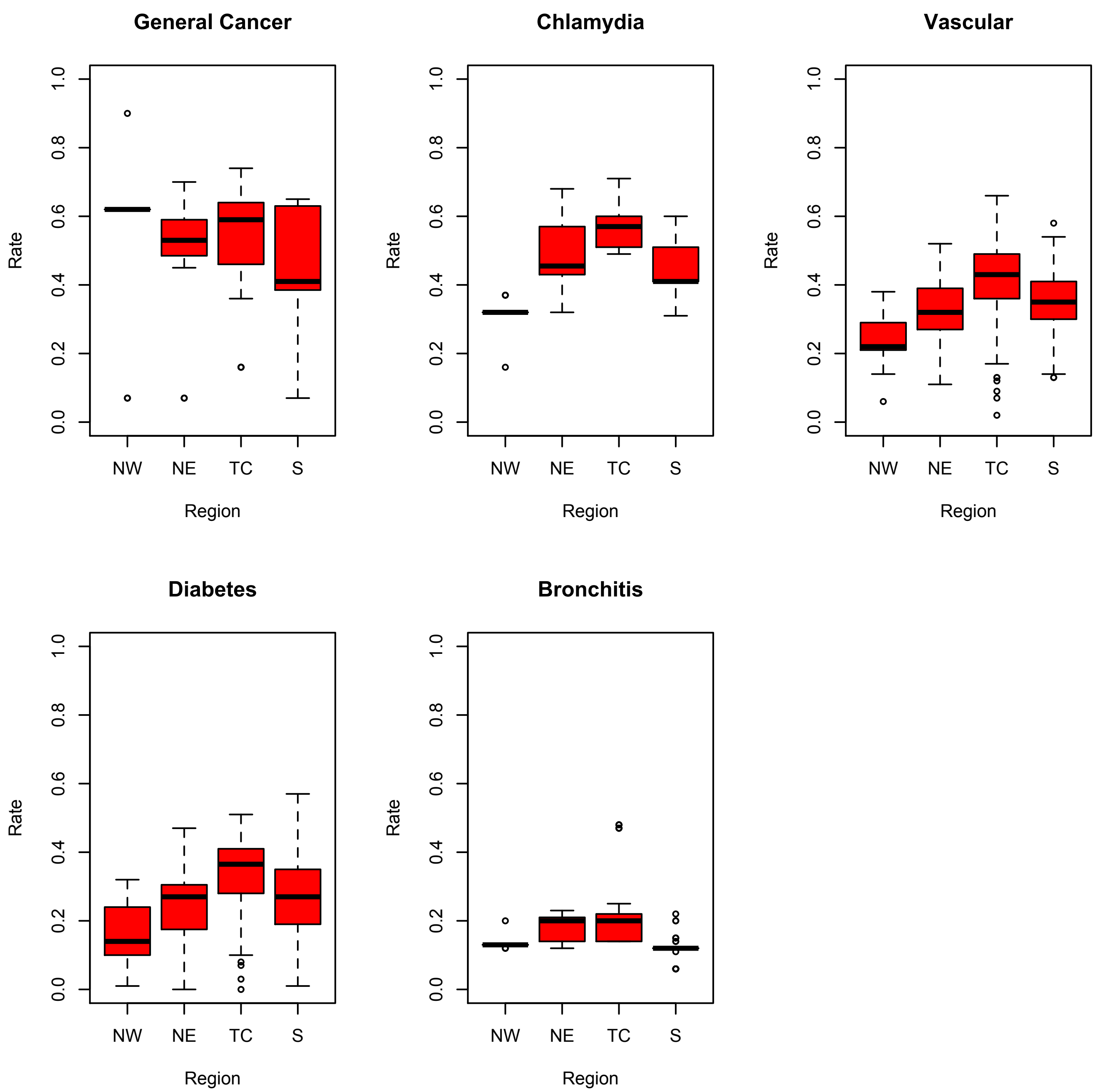

Figure 3: Boxplots of the five procedures with the lowest mean observed rates, broken down by region.

Table 2: The means in each region on each of the observed rate variables rounded to two decimal places.

Measure	Region NW	Region NE	Region TC	Region S
Asthma	.92	.88	.92	.93
Common Cold	.86	.86	.92	.81
Sore Throat	.77	.83	.93	.75
Breast Cancer	.83	.81	.84	.82
Immunization	.84	.81	.85	.77
Cervical Cancer	.78	.78	.82	.76
Colorectal Cancer	.73	.72	.74	.68
Blood Pressure	.66	.70	.74	.60
General Cancer	.54	.51	.56	.43
Chlamydia	.32	.49	.58	.44
Vascular	.23	.32	.41	.35
Diabetes	.16	.25	.33	.26
Bronchitis	.13	.18	.20	.12

Table 3: The means in each region on each of the observed rate variables rounded to two decimal places. Each of the rates was calculated on the complete data for a particular variable.

Measure	Overall Mean	Overall SD	Region NW	Region NE	Region TC	Region S
Asthma	.91	.04	.91	.88	.92	.91
Common Cold	.85	.11	.78	.83	.91	.79
Sore Throat	.83	.14	.67	.83	.90	.79
Breast Cancer	.81	.05	.81	.80	.81	.82
Immunization	.79	.09	.82	.77	.82	.74
Cervical Cancer	.78	.05	.76	.76	.80	.77
Colorectal Cancer	.69	.16	.67	.72	.70	.64
Blood Pressure	.65	.14	.62	.64	.71	.57
General Cancer	.49	.18	.47	.49	.53	.36
Chlamydia	.46	.12	.32	.43	.55	.42
Vascular	.33	.13	.20	.28	.37	.31
Diabetes	.22	.13	.13	.19	.26	.20
Bronchitis	.19	.09	.13	.16	.22	.15

a weighted average of all of the rates, and regions TC and S are *almost* separated by a hyperplane perpendicular to that dimension. However, all but two of the variables included in the initial LDA calculation were aggregated across medical group. After removing the aggregated variables, which was done to construct the other three plots in Figure 4, the regional differences disappeared. The disappearance remained regardless of whether the LDA was calculated on the original 206 clinics (the upper right plot), the 366 clinics who reported vascular and diabetes rates (the lower left plot), or the same 366 clinics after adding in variables capturing the percentage of payments due to Medicare (the lower right plot). No sequence of actual variables (i.e. those that actually varied among all of their observations) could recapture the original, striking differences among regions. I even used the adjusted rates rather than the observed rates, and the comparable plots looked nearly identical.

One might want to quantify the different conclusions reached in the discriminant analyses. A simple approach to quantification is the misclassification rate calculated through leave-one-out-cross-validation. To do this for all of the variable sets, I removed clinic i, estimated LDA coefficients on the remaining clinics, and predicted the region of clinic i based on those coefficients. After repeating this procedure for all of the clinics, I summed the correct classifications and divided by the total number of clinics. Not surprisingly, the misclassification rate associated with the upper-left plot of Figure 4 is approximately .17, whereas the other misclassification rates are much higher (.51 for the upper-right plot, .80 for the bottom-left plot, and .44 for the bottom-right plot). However, the latter three models were based off of fewer variables, which could have contributed to the higher misclassification rates. To address this concern, I performed cross-validation on the full set of rates, including those that did not vary within a medical group, but rather than removing and predicting each clinic separately, I iteratively removed and predicted entire medical groups. Using medical group as the unit of observation, the misclassification rate jumped to .47.

Ultimately, the strength of classification depended on whether I used all of the rates (including those that did not vary within medical group), or only the small subset of rates that varied among all clinics. One indication that the dramatic separation of the original linear discrimination was an artifact of data reporting is that when predicting entire medical groups rather than individual clinics, the misclassification rate increased. However, a perceptive reader might notice that if medical groups were mostly homogenous, then the initial conclusions would not be artifactual. But, if results based off of variables not aggregated over medical group are any indication, clinics are more diverse than originally presumed. Yet, any reservation can only be hypothetical without more data available. Moreover, even if the regions were not different, perhaps there are better ways of similarly clustering the clinics, and perhaps those clusters are more robust to the method of data reporting.

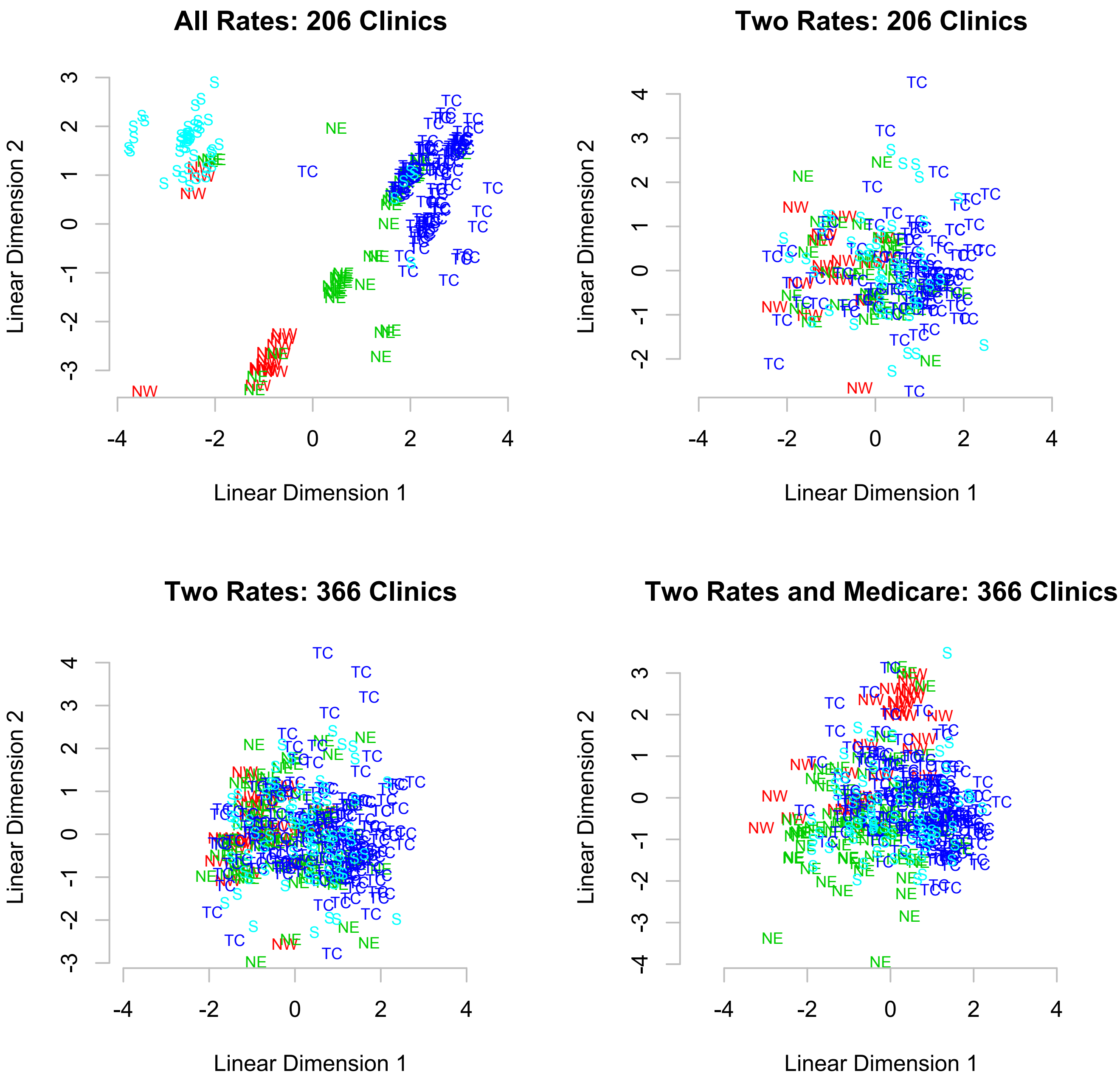

Figure 4: Scatterplots of the first two linear discriminant dimensions predicting region from the observed rate variables. The linear discriminant analysis displayed in the upper left plot was calculated on the 206 completely observed clinics and includes all of the observed rates. The linear discriminant analysis displayed in the upper right plot was calculated on the 206 completely observed clinics and includes only diabetes and vascular observed rates. The linear discriminant analysis displayed in the lower left plot was calculated on the 366 clinics that reported rates for diabetes and vascular and includes only diabetes and vascular observed rates. The linear discriminant analysis displayed in the lower right plot is similar to that displayed in the lower left plot but adds the rate of Medicare payment for both diabetes and vascular procedures.

3 Cluster Analysis

Because all of the clinics within a particular medical group did not vary on most of the rates, any cluster procedure using the original set of clinics should assign clinics to clusters defined by their medical group. Nevertheless, the cluster machinery provides an interesting visualization of stochastic dependence. For instance, I used a model-based cluster algorithm[2] on the reduced set of data with observed rates as the variables. Pretending that I did not know the medical group to which each clinic belonged, I chose the best set of clusters based on a BIC criterion, then iteratively removed outliers, and then reran the cluster algorithm forcing the number of clusters to be fewer than 15. I did the latter for two reasons: (1) there should always be a small number of clusters relative to number of data points; and (2) increasing the maximum number of clusters resulted in the cluster algorithm separating clinics belonging to the same medical group, presumably based on scores to diabetes and vascular observed rates. Ultimately, the algorithm terminated with 14 clusters, all of the same shape, and lying along the coordinate axes. Not surprisingly, the cluster algorithm detected most of the medical groups; however, some of the medical groups were not obvious from the data. For example, all of the algorithms grouped Community University Health Care, Fremont Community Health Services, Open Cities Health Center, and West Side Community Health Services together, even though their medical group labels were not identical. But as in every other medical group, all but two of the observed rates did not vary among those clinics.

Four coordinate projections[3] of the final set of clusters are displayed in Figure 5, and as before, the contrast between the upper left plot and the other three plots is striking. The upper left plot contains the two variables allowed to vary among clinics within the same medical group. Even though some points lie within their cluster ellipse, the ellipses are spread out, and many points do not appear to neatly lie in any ellipse. That is to say, the points in the upper left plot form a standard, linear scatter and not a disparate set of clusters. The other plots indicate that there

[2]Model-based clustering is an implementation of the EM-algorithm, in which the *categories* belonging to each of the objects are unobserved. That is to say, the complete data for clinic i is $(\mathbf{x}_i, \mathbf{z}_i)$, where $\mathbf{x}_i$ is a p-dimensional vector of observations (p is the number of observed variables), and $\mathbf{z}_i$ is a k-dimensional vector of classifications (k is the number of groups) such that

$$z_{ik} = \begin{cases} 1 & \text{if } \mathbf{x}_i \text{ belongs to group k.} \\ 0 & \text{otherwise.} \end{cases}$$

where z_{ik} is drawn from a multinomial distribution with unknown probabilities $\tau_1, \dots, \tau_K$. The E-step (Expectation) of the EM-algorithm estimates $\mathbf{z}_i$ for $i \in \{1, \dots, n\}$, and the M-step (Maximization) maximizes the likelihood with respect to distributional parameters θ_k for $k \in \{1, \dots, K\}$ and class probabilities τ_k for $k \in \{1, \dots, K\}$. To maximize the likelihood, a particular probability distribution must be specified for $\mathbf{x}_i|\theta_k$ (the within group k vector of observations), and it is usually assumed to be multivariate normally distributed [1].

[3]Because the clusters are assumed multivariate normally distributed, each of the clusters is represented by an ellipsoid in multivariate space. The shape, size, eccentricity, and location of each ellipsoid depends on the mean vector and covariance matrix. When projecting the ellipsoid onto the space of two variables, it appears as a two dimensional ellipse [3, pp. 156 – 158].

are bivariate relationships among the other variables; however, those bivariate relationships relate the average of the clinics within a given medical group and therefore, they probably overstate the degree of relationship. Moreover, the four coordinate projections displayed are among the most interesting that I found when sifting through all of the possible pairs of rates. For instance, the relationship between sore throat and cancer is essentially 0; the relationship between two different types of cancer is moderately positive; and the relationship between breast cancer and bronchitis is slightly negative with two outliers. Not all rate variables appear to be positively related, and the extent of relationships might have been exaggerated by the method of data collection.

One might wonder whether there is any residual clustering after taking into consideration medical group. In other words, would removing the redundancy lead to any discernable cluster structure? To answer this question, I aggregated within each of the clusters from the previous step, and I attempted to form model-based clusters using the within-medical-group means. In all cases, using variable sets constructed from: the observed rates; the percentage of Medicare payments; and the percentage of MinnesotaCare payments, the cluster algorithm deemed as many clusters as number of objects. Without within-cluster variability, the clusters are separated enough in multivariate space to prevent additional clustering. As another option, rather than forming model-based clusters, I tried a hierarchical cluster algorithm with Euclidean distance and complete linkage, displayed in Figure 6 (see Gan, Ma, & Wu, 2007 [2]). Unlike the model-based clustering, several of the medical groups appeared to form cluster pairs with those of similar observed rates, and the cluster pairs did not depend on the variable sets in the model. Furthermore, regardless of whether I included Medicare, MinnesotaCare, or solely the observed rates, the locations of most of the branches were in the same place. Thus, the clustering found an overall, general difference in the rates, but it is not obvious whether the cluster algorithm found anything more complicated.

Figures 7 – 9 display boxplots of observed rates for clinic performance on all of the conditions broken down by hierarchical cluster. To determine which medical groups clustered together, I used the upper left dendrogram of Figure 6, constructed solely from the observed rates. Cluster 1 consists of Mayo Health, Affiliated, and Mayo Clinic. Cluster 2 consists of MeritCare, Buffalo, St. Mary's, and Allina 1. Cluster 3 consists of Fairview, HealthEast, Aspen, Park Nicolett, HealthPartners, Allina 2, and St. Luke's. Cluster 4 consists of "Cluster 5" and Avera. Cluster 5 consists of "Cluster 9" and HCMC. I determined which clinics belonged to a given hierarchical cluster based on proximity in the dendrogram, but I made sure that each cluster consisted of at least two medical groups. Furthermore, unlike Figures 1 – 3, I took the mean rate within a medical group as the unit of observation, and I did not weight the rates by the number of clinics contributing to them. Clusters 4 and 5 tend to have lower rates for most of the procedures, whereas cluster 3 tends to have higher overall rates. Moreover, several of the procedures appear to separate one or two clusters from the remaining clusters. For instance, cluster 4 has lower observed rates on the common cold and sore throat than the other clusters, and cluster 5 has higher rates

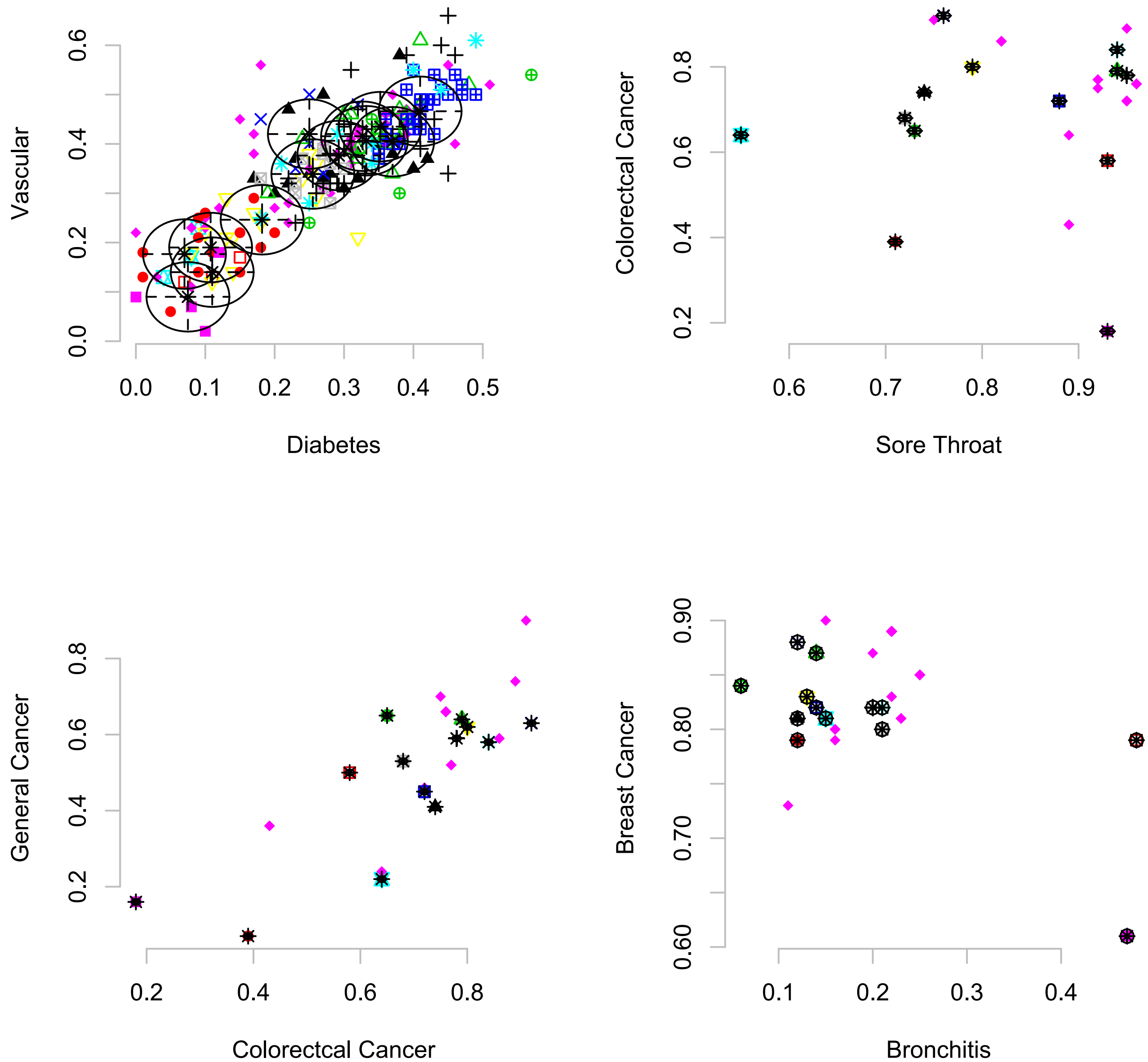

Figure 5: Coordinate projections of fourteen model-based clusters along specified coordinate axes after restricting the number of clusters to fewer than 15. The color and shape of points identifies clinics belonging to the same cluster, and the ellipse represents the shape, size, and orientation of each cluster containing the clinics. Unclusterable clinics are represented by pink diamonds, and they were iteratively removed prior to determining the final set of clusters.

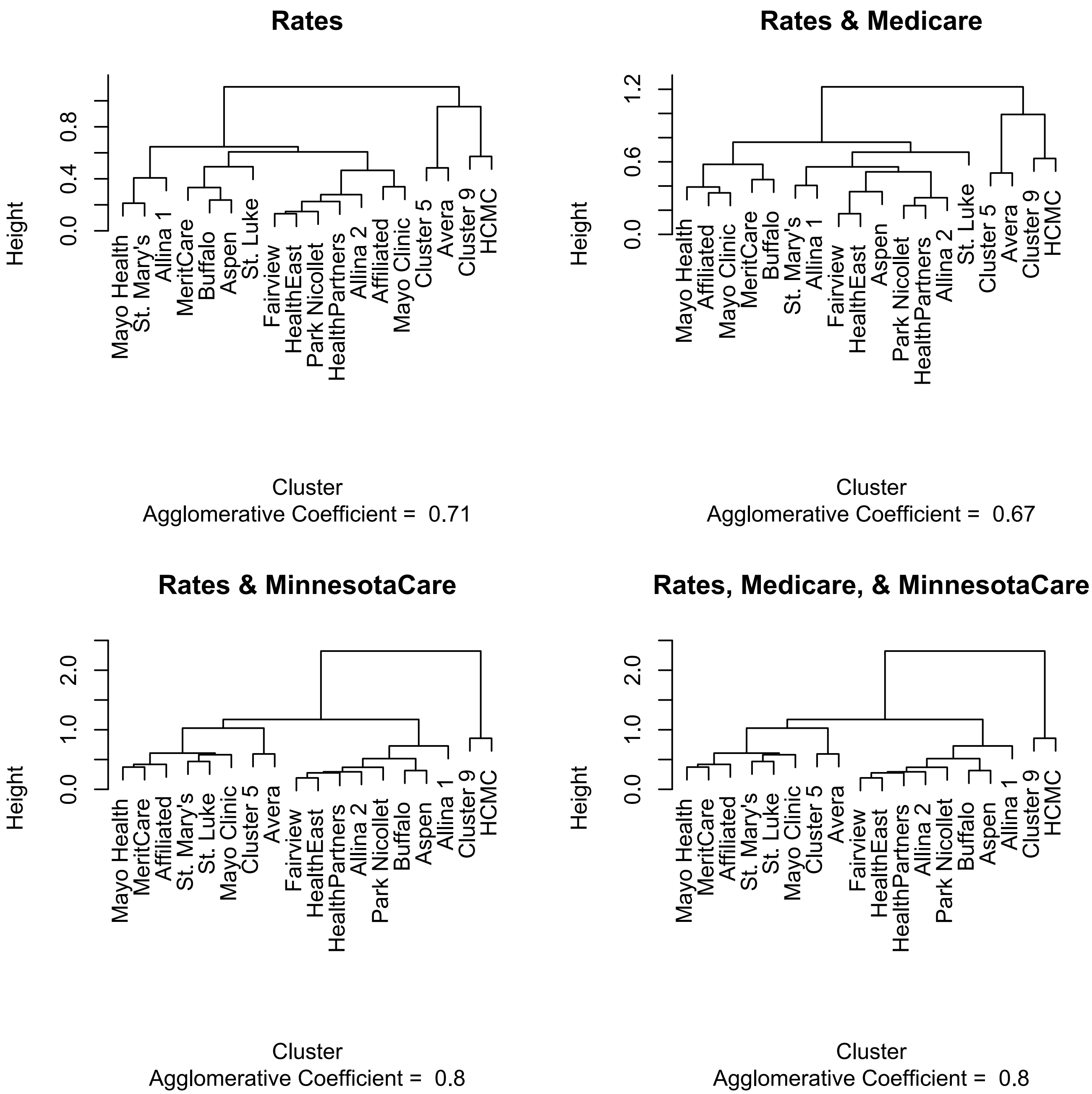

Figure 6: Dendrograms of hierarchical clusters based on the aggregated dataset. Each object represents a medical group, and clinic rates are averaged over medical group. The upper left plot includes all of the observed rates; the upper right plot includes observed rates and the percentage of Medicare payments; the lower left plot includes observed rates and the percentage of MinnesotaCare payments; and the lower right plot includes observed rates and both of the payment variables.

on bronchitis than the other clusters. However, because all of the clusters are based on so few observations, and because the observations consist of means within a medical group, it is difficult to draw more suitable conclusions. To understand the loss of information, compare the boxplots of diabetes and vascular observed rates from Figures 1 – 3, where each clinic had a unique rate, to Figures 7 – 9, where the rates were aggregated within medical group. The latter boxplots are similar in appearance to boxplots of the redundant variables, suggesting stronger relationships between cluster and rate than is probably the case.

To determine what *might* have happened had individual rates been reported for clinics within the same medical group, I formed clusters based solely on diabetes and vascular. Rather than only including the observed rates, I also included the percentage of Medicare payments.[4] I formed the clusters in the same way as I did before, starting with an initial estimate, iteratively removing outliers, and choosing the structure with the highest BIC. The form of the within-cluster variance matrix is similar to that of before, being of the same size and shape across all of the clusters. However, the interpretations drawn from the coordinate projections are illuminating, as shown in Figure 10. Now, there appears to be a distinct relationship between observed rate, Medicare payment, and cluster membership. As shown in the upper-left plot, the green (circle) cluster has higher observed rates on both of the variables, whereas the black (triangle) and red (square) are nearly equal in performance. However, the upper right and lower left plots clearly show that the black cluster has a high percentage of Medicare payments whereas the red cluster has a low percentage of Medicare payments. And as is obvious from the lower right plot, the cluster with the highest observed rates is in the middle in terms of Medicare payments. The means and standard deviations of the observed rates, Medicare payments, and MinnesotaCare payments within each of the clusters is displayed in Table 4. Save for MinnesotaCare, the standard deviations are nearly identical within each of the clusters, which is to be expected given the cluster structure. Furthermore, cluster 2 (the red cluster) has the lowest percentage of Medicare payments, but it appears to have the highest percentage of MinnesotaCare payments, on average. However, a small proportion of cluster 2 clinics are heavily influencing the average, which explains the large within-cluster standard deviation on MinnesotaCare variables for cluster 2.

Unlike earlier assumed, the resulting clusters were *not* homogenous with respect to medical group. For instance, HealthEast, one of the best performing medical groups with respect to all of the rates, had seven clinics included in cluster 3 (the green cluster) but two clinics contained in cluster 1 (the black cluster). Furthermore, the rates for those two clinics on were much lower than the corresponding rates for the other seven variables. Thus, even though most of the clinics within a given medical group clustered together, it would be presumptuous to assume that the mean rate across medical group should be taken as a substitute for the performance of individual clinics.

[4] I also tried estimating the clusters by adding in the MinnesotaCare variable, but few clinics have a majority of their payments by MinnesotaCare, so that the few outliers on MinnesotaCare had undue influence.

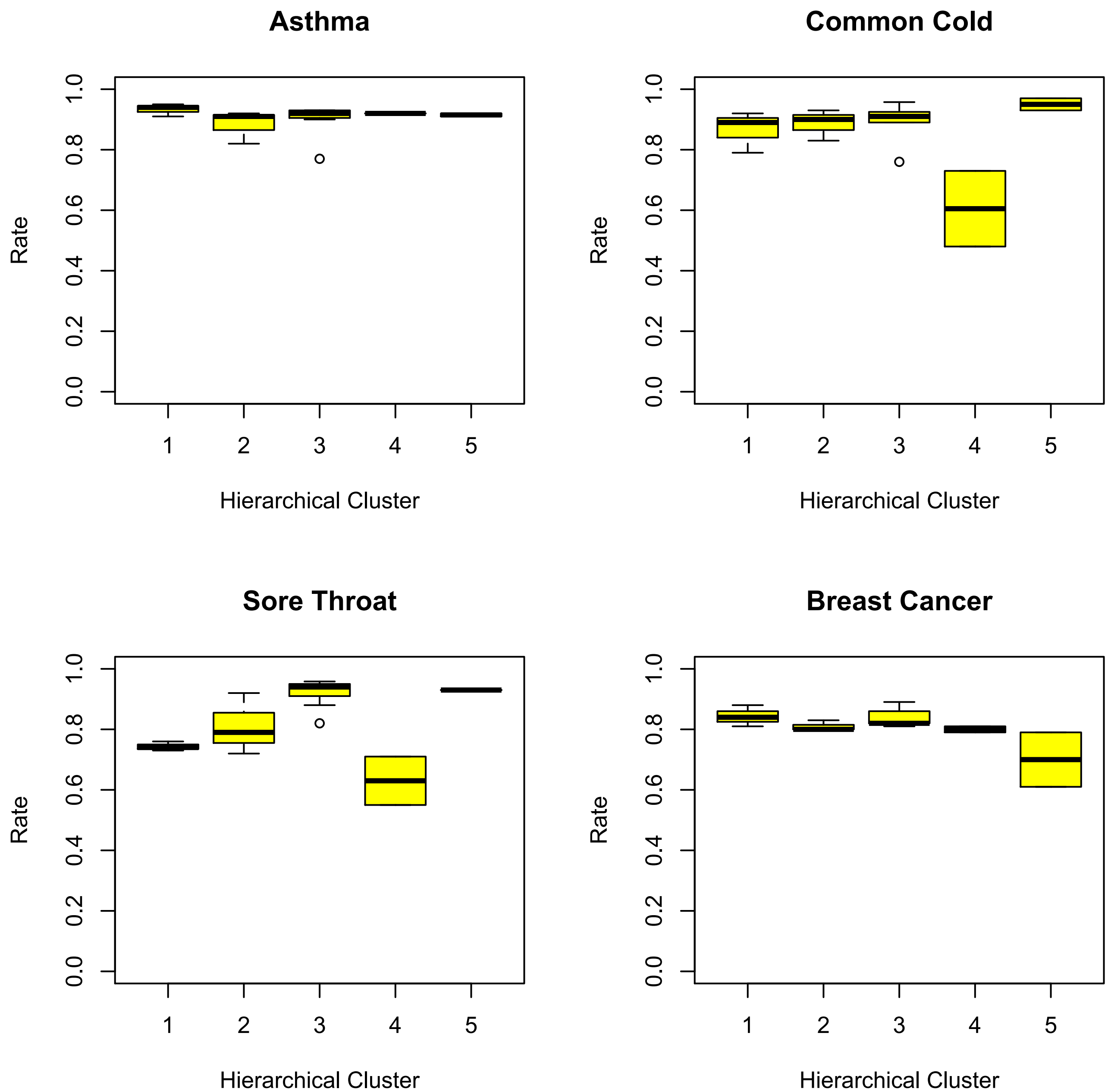

Figure 7: Boxplots of the four procedures with the highest mean observed rates, broken down by hierarchical cluster. Cluster 1 consists of Mayo Health, Affiliated, and Mayo Clinic. Cluster 2 consists of MeritCare, Buffalo, St. Mary's, and Allina 1. Cluster 3 consists of Fairview, HealthEast, Aspen, Park Nicolett, HealthPartners, Allina 2, and St. Luke's. Cluster 4 consists of "Cluster 5" and Avera. Cluster 5 consists of "Cluster 9" and HCMC.

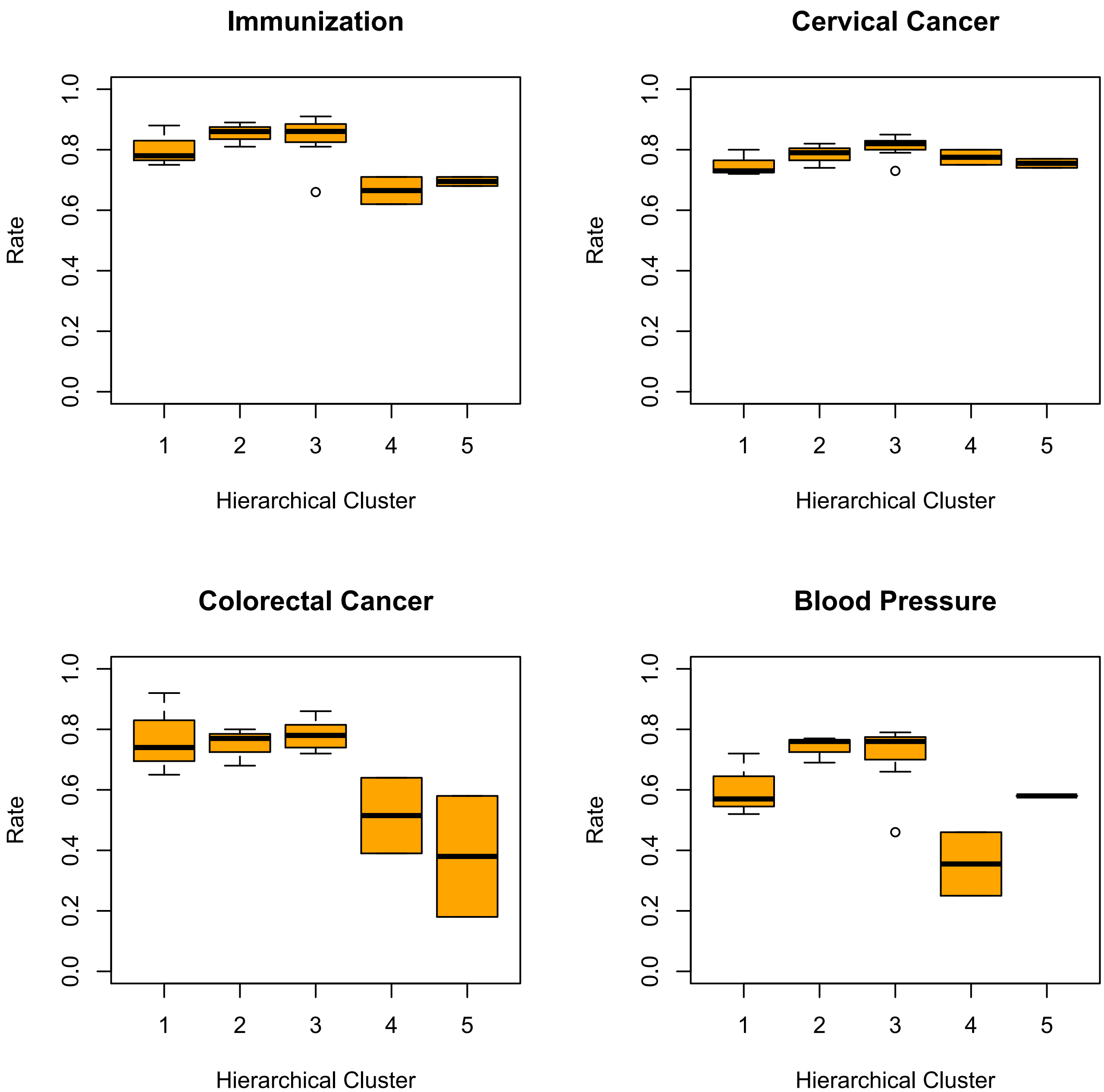

Figure 8: Boxplots of the four procedures with the middle mean observed rates, broken down by hierarchical cluster. Cluster 1 consists of Mayo Health, Affiliated, and Mayo Clinic. Cluster 2 consists of MeritCare, Buffalo, St. Mary's, and Allina 1. Cluster 3 consists of Fairview, HealthEast, Aspen, Park Nicolett, HealthPartners, Allina 2, and St. Luke's. Cluster 4 consists of "Cluster 5" and Avera. Cluster 5 consists of "Cluster 9" and HCMC.

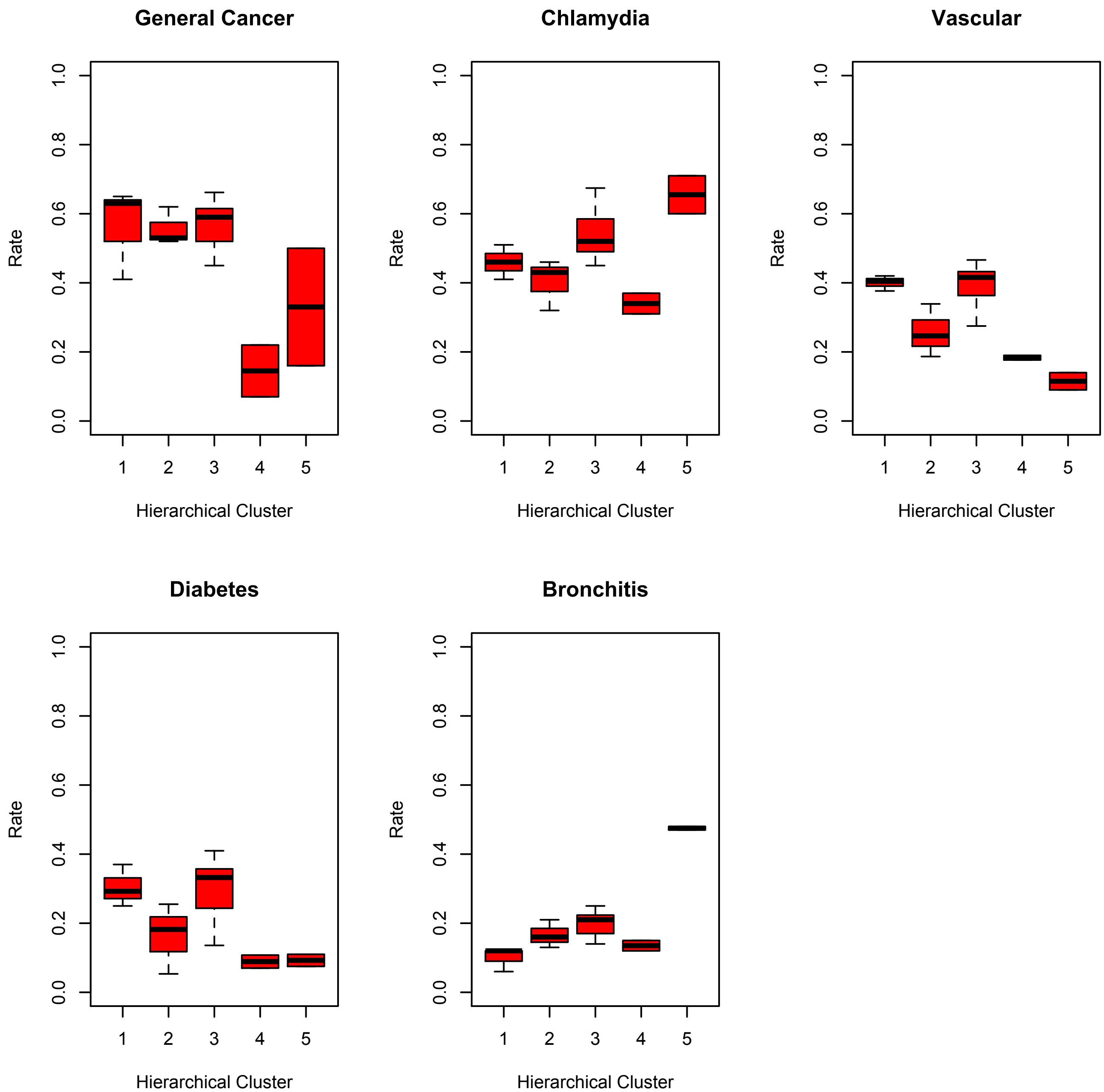

Figure 9: Boxplots of the four procedures with the lowest mean observed rates, broken down by hierarchical cluster. Cluster 1 consists of Mayo Health, Affiliated, and Mayo Clinic. Cluster 2 consists of MeritCare, Buffalo, St. Mary's, and Allina 1. Cluster 3 consists of Fairview, HealthEast, Aspen, Park Nicolett, HealthPartners, Allina 2, and St. Luke's. Cluster 4 consists of "Cluster 5" and Avera. Cluster 5 consists of "Cluster 9" and HCMC.

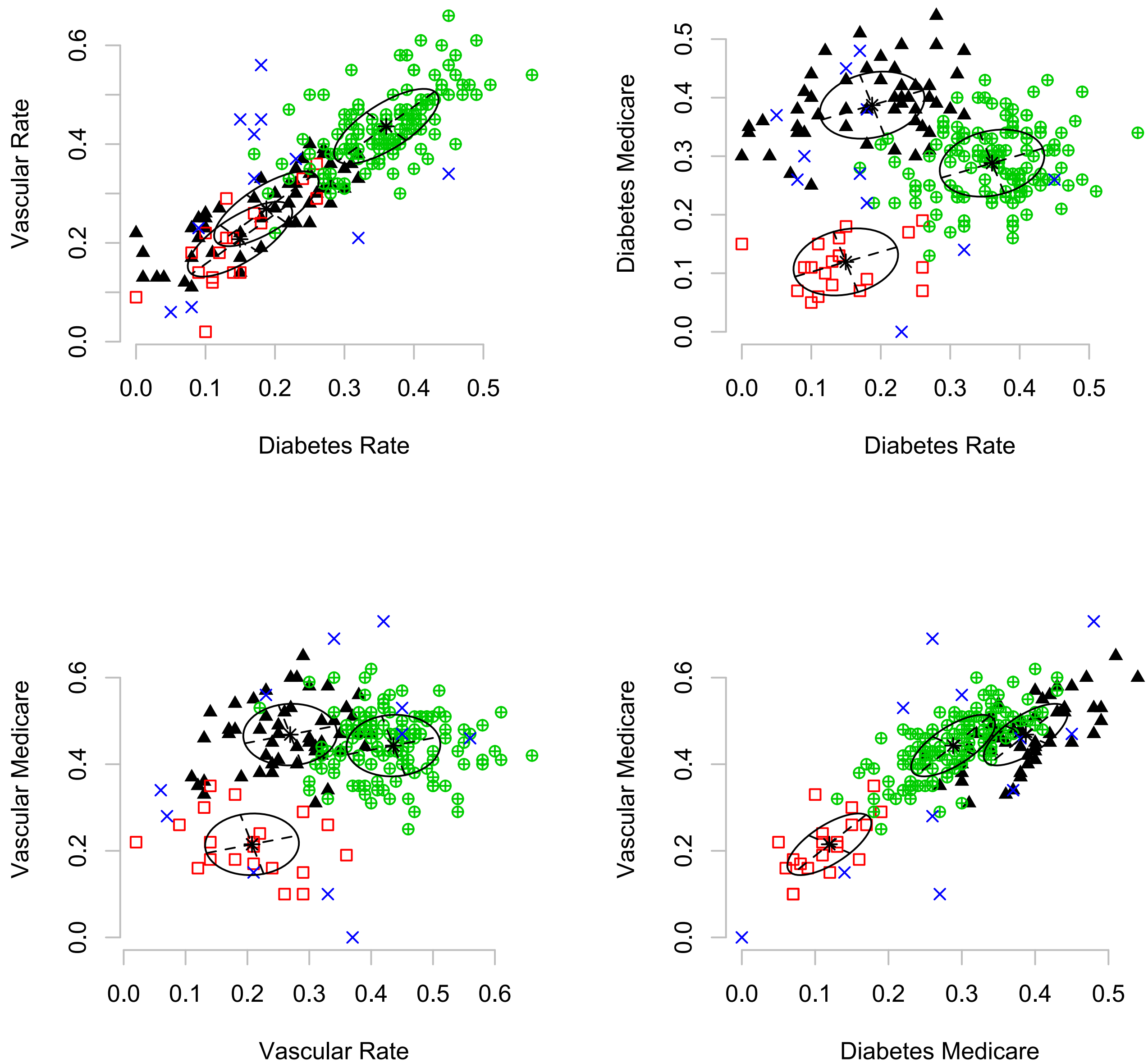

Figure 10: Coordinate projections of four model-based clusters along the coordinate axes The color and shape of points identifies clinics belonging to the same cluster, and the ellipse represents the shape, size, and orientation of each cluster containing the clinics. Unclusterable clinics are represented by blue x's, and they were iteratively removed prior to determining the final set of clusters.

Table 4: The means and standard deviations in each of the clusters on some of the predictors that varied amongst clinics within a given medical group. Only observed rates and Medicare payments determined the clusters, but other variables are displayed for comparison.

Cluster Means

Cluster	Diabetes Rate	Diabetes Medicare	Diabetes MNCare	Vascular Rate	Vascular Medicare	Vascular MNCare
Cluster 1	.18	.38	.15	.27	.46	.10
Cluster 2	.15	.12	.30	.20	.21	.21
Cluster 3	.36	.29	.13	.44	.44	.08
Unclusterable	.19	.28	.15	.32	.39	.48

Cluster Standard Deviations

Cluster	Diabetes Rate	Diabetes Medicare	Diabetes MNCare	Vascular Rate	Vascular Medicare	Vascular MNCare
Cluster 1	.09	.06	.09	.08	.08	.09
Cluster 2	.07	.04	.24	.09	.07	.19
Cluster 3	.07	.06	.07	.08	.07	.04
Unclusterable	.11	.14	.17	.16	.24	.18

4 Conclusion

Unfortunately, the well formed clusters was an artifact of sampling methodology and not necessarily an attribute of clinics. The within cluster variability of diabetes and vascular observed rates of performance (the only two variables allowed to vary within a medical group) is high enough that it would be a mistake to assume that all clinics within one medical group are equal on performance. Furthermore, by listing each clinic separately *apart* from their medical group, the authors of the health summary might deceive those who are searching for honest information about a specific clinic. I understand that a particular clinic might not have enough patients undergoing a specific procedure to allow for specific rates reported for that clinic. However, prospective patients deserve a clearer description of the tables and a better explanation of the use of confidence intervals. Perhaps, rather than reporting *confidence intervals*, the Minnesota Department of Health could report *variance estimates* for the rate parameters within a given medical group. When a researcher publishes a confidence interval and rate next to a particular clinic, most educated people would assume that the confidence interval and rate apply to the specific clinic and not the medical group to which the clinic belongs. Because each clinic might have a *different* population rate parameter for a given procedure, it is misleading to assume that the mean rate of a medical group is identical to the mean rate of a given clinic. Moreover, it is statistically criminal to imply that a confidence interval calculated on the mean rate of a medical group is the same as the confidence interval for a given clinic. The latter would need a prediction interval or an indication of the differences between a given clinic and the mean of its medical group.

Based on the analysis of the two clinic procedures allowed to vary among clinics within medical group, there might be a more interesting cluster structure obscured by methodological peculiarities. For those interested in deciding which clinic to attend, there is little else to recommend than to use the rates associated with its medical group, and displayed in Table 5. Yet, be warned. Just because a medical group has a high rate, the associated clinics might not all be uniform with respect to performance. In the future, a more detailed description of clinic performance and/or an assessment of the uncertainty of individual clinics within a medical group would allow for thorough analyses and more confident recommendations.

References

[1] Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, *97*, 611 – 631.

[2] Gan, G., Ma, C., & Wu, J. (2007). *Data Clustering Theory, Algorithms, and Applications*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

[3] Härdle, W., & Simar, L. (2007). *Applied Multivariate Statistical Analysis*. New York, NY: Springer-Verlag.

[4] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*, 18 – 22.

[5] Minnesota Department of Health. (2011). Physician Clinic Risk Adjustment Fact Sheet. Retrieved February 16, 2011 from http://www.health.state.mn.us/healthreform/measurement/report/RiskAdjustmentFactSheet_Nov172010.pdf.

[6] Minnesota Statutes ch. 62U, §02 (2010).

[7] Pitches, D. W., Mohammed, M. A., & Lilford, R. J. (2007). What is the empirical evidence that hospitals with higher-risk adjusted mortality rate provide poorer quality care? A systematic review of the literature. *BMC Health Services Research*, *7*. Retrieved February 16, 2011 from http://www.biomedcentral.com/content/pdf/1472-6963-7-91.pdf.

[8] Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147 – 177.

[9] Schafer, J. L. (2010). mix: Estimation/multiple imputation for mixed categorical and continuous data. R package version 1.0-8. URL http://CRAN.R-project.org/package=mix.

Table 5: The means in each cluster on each of the observed rate variables rounded to two decimal places. The variables are ordered from left to right from highest overall rate to lowest overall rate whereas the medical groups are ordered from top to bottom from lowest overall rate to highest overall rate. Misc contains the clinics that did not fit in the other medical groups.

		Procedures of Highest Overall Rates					
Medical Grp.	Overall Mean	Asthma	UTI	Pharyngitis	Breast	CHL	Cervical
Avera	.46	.92	**.48**	**.55**	.81	**.62**	.80
Cluster 5	.49	.92	.73	.71	.79	.71	.75
Cluster 9	.54	.91	.93	.93	**.61**	.68	.74
St. Luke's	.58	.77	.76	.82	.81	.66	.73
Mayo Health	.60	.91	.79	.74	.81	.75	**.72**
St. Mary's Duluth	.61	**.82**	.83	.72	.80	.81	.74
Allina 1	.62	.91	.91	.90	.83	.73	.81
MeritCare	.62	.91	.90	.79	.83	.89	.79
Buffalo	.63	.92	.93	.92	.80	.86	.82
HCMC	.63	.92	**.97**	.93	.79	.71	.77
Affiliated CMC	.64	**.95**	.89	.73	.84	.88	.80
Mayo Clinic	.64	.94	.92	.76	.88	.78	.73
Aspen	.64	.91	.91	.95	.85	.88	.82
Allina 2	.67	.92	.92	.88	.82	.81	.81
Fairview	.68	.93	.89	.95	.82	.84	.82
HealthEast	.69	.90	.93	.94	.82	**.91**	.79
Park Nicollet	.69	.93	.89	.94	.87	.89	**.85**
HealthPartners	.72	.93	.96	**.96**	**.89**	.86	.84
Misc.	.62	.95	.83	.86	.80	.68	.78
Overall Mean	.64	.91	.88	.86	.83	.82	.79

		Procedures of Lowest Overall Rates					
Medical Grp.	Colorectal	Blood Pres.	Cancer	Chlamydia	Vasc.	Diabetes	Bronchitis
Avera	.64	**.25**	.22	**.31**	.18	.07	.15
Cluster 5	.39	.46	**.07**	.37	.19	.11	.12
Cluster 9	**.18**	.58	.16	.60	**.09**	.07	.47
St. Luke's	.86	.46	.59	.45	.28	.16	.23
Mayo Health	.74	.72	.41	.41	.38	.29	.12
St. Mary's Duluth	.68	.76	.53	.43	.34	.26	.21
Allina 1	.43	.75	.36	.51	.38	.31	.22
MeritCare	.80	.69	.62	.32	.25	.18	.13
Buffalo	.77	.77	.52	.46	.19	**.05**	.16
HCMC	.58	.58	.50	**.71**	.14	.11	**.48**
Affiliated CMC	.65	.57	.65	.51	.40	.37	**.06**
Mayo Clinic	**.92**	.52	.63	.46	.42	.25	.12
Aspen	.72	.66	.46	.49	.32	.14	.25
Allina 2	.72	.76	.45	.60	**.47**	**.41**	.14
Fairview	.78	.74	.59	.57	.41	.33	.20
HealthEast	.84	**.79**	.58	.52	.42	.33	.21
Park Nicollet	.79	.78	.64	.49	.44	.35	.14
HealthPartners	.76	.77	**.66**	.67	.43	.36	.22
Misc.	.81	.65	.63	.36	.30	.20	.16
Overall Mean	.72	.69	.52	.50	.36	.29	.17

[10] Schiff, J., Adams, P., & Leitz, S. (2008, December). *Health care reform update [Power Point Slides]*. Accessed April 26, 2011 from: www.commissions.leg.state.mn.us/lchca/healthcarereform.ppt.

[11] Su, Y-S., Gelman, A., Hill, J., & Yajima M. (2011). Multiple imputation with diagnostics (mi) in R: Opening windows into the black box. *Journal of Statistical Software*. Accessed February 28, 2011 from: http://www.jstatsoft.org.

[12] Thomas, J. W., & Hofer, T. P. (1999). Accuracy of risk-adjusted mortality rate as a measure of hospital quality of care. *Medical Care*, *37*, 83 – 92.