Paper 115-27

Using SAS® Enterprise Miner™ for Data Quality Monitoring in the Veterans Health Administration's External Peer Review Program

James H. Forsythe, Ph.D., West Virginia Medical Institute, Charleston, WV
A.B. Billings, M.S., West Virginia University, Morgantown, WV
Jonathan B. Perlin, M.D., Ph.D., Veterans Health Administration, Washington, DC
John Brehm, M.D., West Virginia Medical Institute, Charleston, WV
Scott A. Miller, M.A., West Virginia Medical Institute, Charleston, WV

ABSTRACT

In support of the Veterans Health Administration's External Peer Review Program (EPRP) the West Virginia Medical Institute performed over 350,000 medical record reviews during FY 2001. Over 100 highly trained abstractors (QICs) conduct record reviews in a variety of settings that include 170 VHA hospitals and more than 500 outpatient clinics. To help assure the quality of the medical record abstractions, WVMI monitors abstractor performance and the validity and reliability of the medical record data. WVMI employs a scientifically rigorous multi-method approach for assessing data quality and (of note), uses SAS® Enterprise Miner™ for improving the precision and accuracy of assessing the abstracted data. This paper describes the use of SAS® EM™ and how applying the software has aided WVMI in determining the source of unexpected or "anomalous" abstractions. These advancements greatly strengthen the "bottom line" value of third party abstractions for assessing clinical performance. Our advancements in the use of SAS® EM™ for assessing abstractor performance are now being developed for use on a similar contract with the Department of Defense medical treatment facilities.

INTRODUCTION

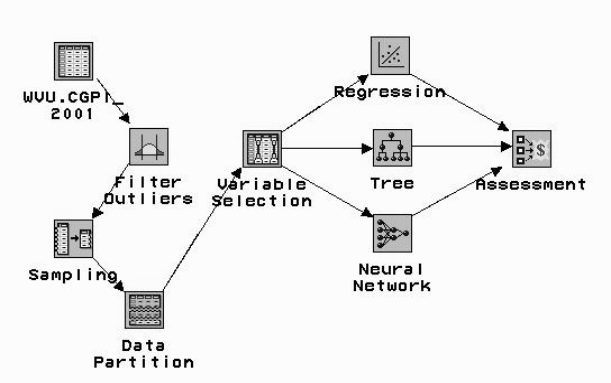
Under the Veterans Health Administration External Peer Review Program (EPRP) program, the West Virginia Medical Institute conducts over 350,000 medical record reviews in 170 hospitals and over 500 clinics. As part of this work, WVMI monitors abstractor performance and conducts data quality assessments to evaluate the validity and reliability of abstracted medical record data. WVMI employs a scientifically rigorous multi-method approach for assessing data quality and, over the life of the program, has implemented state-of-the-art techniques for detecting irregularities in abstracted data. The variety of methods and techniques for data monitoring has begun to emerge as a comprehensive model for detecting and correcting data problems. The advances in data quality control and improvement began in earnest in 1999. Of note, as the complexities of assessing guideline performance have increased, WVMI has introduced improved methods for assessing the precision and accuracy of abstracted data. These parallel advancements (increased guideline measurement complexity and methods of data quality assessments) are complementary and greatly strengthen the "bottom line" value of the data being used to assess clinical performance. For example, WVMI has supplemented principal data quality activities such as inter-rater reliability assessments and retrospective, ad hoc analyses with new techniques such as computer-aided "real-time" data screening1 Further, WVMI in concert with EPRP leadership, have promoted a scientific approach to data quality control and improvement such that, where methodologically constrained (as with the field-based interrater assessments), alternative approaches for validating performance have been successfully implemented.[1]

SAS® EM™ & DATA QUALITY MONITORING

SAS® Institute has developed and markets data mining software called Enterprise Miner™. SAS® Enterprise Miner™ is an integrated software system that provides end-to end capabilities for data mining. SAS® EM™ provides a graphical user interface (GUI) for user-friendly implementation of the data mining process. The GUI is designed so that a user with little statistical experience can navigate through the data mining methodology, while allowing a user with substantial statistical expertise to fine-tune the analytic process. Intelligent use of data mining methods requires expertise in three areas. The *domain expert* understands the particulars of the problem including relevant background knowledge of the problem, the context in which the problem arises, and the appropriate terminology. The *data expert* understands the structure, size and format of the data. The *analytic expert* understands the capabilities and limitations of the methods that may be appropriate to the problem. The embodiment of the required expertise may take one, two, three, or more people.

We have found several advantages to using SAS® EM™ for purposes of data quality modeling. First, SAS® EM™ allows easy user control of the process flow. Current data mining software, e.g. SAS® EM, utilizes GUIs that allow the user to create and control analyses in terms of a "flow diagram" (See Figure 1). The user may focus on the general aspects of the analysis, while minimizing the effort of writing programs to perform the analysis. SAS® EM™ allows the user to generate complicated models of various types (regression, neural nets, and decision trees) with minimal effort. SAS® EM™ provides methods for quickly and easily comparing qualitatively distinct models. For example, regression models, neural network models, and decision tree models can be assessed and compared. SAS® EM™ provides methods for storing and using models to score (predicting the value of a target variable) new observations. SAS® EM™ allows the use of existing code to perform data preparation and customized analysis. New code can also be generated.

FIGURE 1 SAS® EM™ VISUAL PROGRAMMING INTERFACE



SAS® EM™ contains advanced visualization tools to aid in the interpretation and understanding of results. SAS® EM™ provides a great amount of flexibility to the user via the opportunity to control parameters and options used in the analysis. The user can control the structure of the analysis, can choose various optional methods for generating models, and can modify default values of parameters. SAS® EM™ can reduce the amount of time and effort required to perform even simple types of analyses. Corresponding to each node and the connections between nodes, SAS® EM™ automatically

generates SAS® statements to perform the task. The generated code does not have to be programmed by the user. The code generated by SAS® EM™ is also available to the user for custom modification. SAS® EM™ produces output in both tabular form and graphical form. The graphical output facilitates interpretation of results and is typically of much higher quality than the graphical output from base SAS® or SAS/Stat®.

RESULTS

Several analyses of key performance indicators (chosen by WVMI) have been conducted. For the first two applications described below, WVMI used a large set of data (over 126,000 records) from the EPRP data. WVMI used a subset of the data (approximately 20,000 records). Each data set contained 267 variables per record. The subset used by WVMI was generated using random sampling. The results of selected analyses are discussed below.

PREDICTIVE MODELING

SAS® EM™ was used to construct predictive models of various dependent variables (determined by WVMI) as a function of other variables on which data were collected. The dependent variables that were modeled include blood pressure measurements, colonoscopy refusal rates, and flu vaccination rates, among others.

For each dependent variable, three models were generated: a logistic regression model, a neural network model, and a decision tree model. The initial runs were performed using the default estimation methods and default values of parameters. The data were partitioned into three subsets: a training data set containing 37,904 records, a validation dataset containing 37,904 records, and a test data set containing 50,540 records. SAS® EM's "Variable Selection" node was executed in order to eliminate from further consideration those variables that showed little or no association with the dependent variable "colon refusal". (For data sets containing a large number of variables, the "Variable Selection" step can help to reduce processing time required to complete each of the subsequent modeling steps.)

COLONOSCOPY REFUSAL RATES

SAS® EM™ was used to construct a predictive model of the dependent variable "colon refusal" as a function of other variables on which data were collected. The variable "colon refusal" indicates whether or not a patient refused colon screening.

The results of the initial runs, using the default modeling options and default parameter values, were neither useful nor informative. On subsequent runs we changed some of the modeling options and default parameter values. The logistic regression modeling was performed using a forward selection method with an entry level of 0.2, and the variable "colosig" was manually eliminated form the set of independent variables.

The forward selection procedure identified 27 independent variables to be included in the model. Eighteen of the variables selected had associated p-values less than 0.0001. The remaining nine variables had p-values between 0.0132 and 0.1476. No other variables met the 0.2 selection entry criterion. After selecting the variables that met the entry criterion, SAS® EM™ used its default criterion to select a more parsimonious model. The selected model used three independent variables. This model had misclassification rates of 0.019877 for the training data set, 0.02825 for the validation data set, and 0.02736 for the test data set.

The second model fitted to the data was a neural network multiperceptron model using one hidden layer. The neural network modeling of the dependent variable colon_refuse resulted in the selection of 29 input variables. The neural network model had misclassification rates of 0.02103 for the training data set, 0.02248 for the validation data set, and 0.02074 for the test data set.

The third method used to model "colon_refuse" was a classification tree model. Three independent variables were selected to construct the classification tree. The classification tree model had misclassification rates of 0.0233 for the training data set, 0.0265 for the validation data set, and 0.0247 for the test data set.

When used strictly for predictive purposes the neural network seemed to exhibit the best performance of three fitted models. But all of the three fitted models performed well, with very low misclassification rates. An examination of the groupings made by SAS® EM™ in the "Variable Selection" step could lead to further investigation. As an example, the SAS® EM™ grouped variable G_vamc partitioned the VAMCs (Veterans Administration Medical Center) into 8 groups. VAMCs that appear in a group may have similar performance with respect to colon_refuse. Alternately, VAMCs that appeared grouped together may have few similarities, and the groupings may simply be the result of SAS® EM's attempts to maximize Chi-square for variable selection. Further exploration, e.g., distributional analysis or clustering may be appropriate.

TERMINAL DIGIT BLOOD PRESSURE = 0

SAS® EM™ was used to construct a predictive model of the dependent variable "both_bp1" as a function of other variables on which data were collected. The variable both_bp1 indicates whether or not both systolic and diastolic blood pressure readings ended with terminal digit 0.

The logistic regression modeling was performed using a forward selection method with an entry level of 0.1, and the variables bp1, bp2, bp1_2, bp2_2, bp_d_term1, and bp_s_term1were manually eliminated form the set of independent variables (representing the terminal digits of two measurements of systolic and diastolic blood pressures).

The forward selection procedure identified 39 independent variables to be included in the model. Thirty-seven of the variables selected had associated p-values less than 0.0001. The remaining variables had p-values between 0.000 and 0.003. No other variables met the 0.1 selection entry criterion. After selecting the variables that met the entry criterion, SAS® EM™ used its default criterion (VMISC) to select a more parsimonious model. The selected model included only the intercept term. This model had misclassification rates of 0.0983 for the training data set, 0.0997 for the validation data set, and 0.0969 for the test data set.

The second model fitted to the data was a neural network multiperceptron model using one hidden layer. The neural network modeling of the dependent variable colon_refuse resulted in the selection of 29 input variables. The neural network model had misclassification rates of 0.0983 for the training data set, 0.0996 for the validation data set, and 0.0996 for the test data set.

The third method used to model both_bp1 was a classification tree model. Three independent variables were selected to construct the classification tree. The classification tree model had misclassification rates of 0.091 for the training data set, 0.096 for the validation data set, and 0.098 for the test data set.

It should be noted that G_vamc seemed to play an important role in the neural network model and the classification model. Also G_vamc was the first variable selected in the regression model. An examination of the groupings made by SAS® EM™ in the "Variable Selection" step could lead to further investigation. The SAS® EM™ grouped variable G_vamc partitioned the VAMCs into 8 groups. VAMCs that appear in a group may have similar performance with respect to both_bp1. Further exploration, e.g., distributional analysis or clustering may be appropriate.

FLU VACCINATION REFUSAL RATES

SAS® EM™ was used to construct a predictive model of the dependent variable "fluvac_refuse" as a function of other variables on which data was collected. The variable fluvac_refuse indicates whether or not the patient refused flu vaccination.

The logistic regression modeling was performed using a forward

selection method with an entry level of 0.1, and the variables refusal_n, refusal_d, refuse, refusal_total, and fluvacdt were manually eliminated form the set of independent variables.

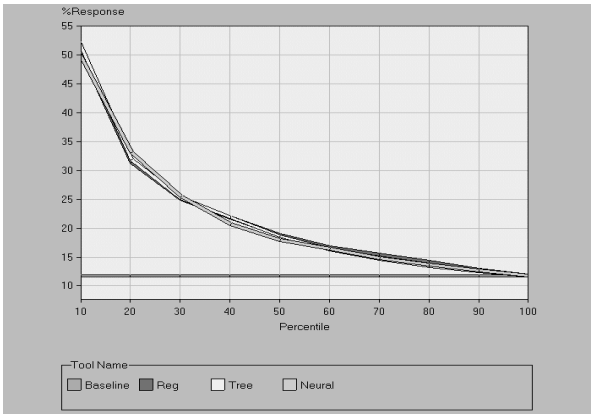
The forward selection procedure identified 4 independent variables to be included in the model. Three of the variables selected had associated p-values less than 0.0001. The remaining variable had p-value between 0.049. No other variables met the 0.1 selection entry criterion. This model had misclassification rates of 0.119 for the training data set, 0.117 for the validation data set, and 0.095 for the test data set.

The second model fitted to the data was a neural network multiperceptron model using one hidden layer. The neural network modeling of the dependent variable fluvac_refuse resulted in the selection of two input variables. The neural network model had misclassification rates of 0.0917 for the training data set, 0.0867 for the validation data set, and 0.0762 for the test data set.

The third method used to model fluvac_refuse was a classification tree model. SAS® EM™ selected three variables to construct the classification tree model. The classification tree model had misclassification rates of 0.0413 for the training data set, 0.1474 for the validation data set, and 0.1442 for the test data set.

When used strictly for predictive purposes there was little difference in the predictive power of the three fitted models (Figure 2 below). In as much as this was an exploratory application of SAS® EM, our next steps are to examine the

FIGURE 2. LIFT CHART OF MODEL PERFORMANCE



groupings made by SAS® EM™ in the "Variable Selection" step. Also, further exploration, e.g., distributional analysis or clustering may be appropriate. The utility of SAS® EM™ for predictive modeling within the context of assessing data quality and QIC performance will be greatly limited by the changes in variables surrounding abstraction activities (e.g., changes in instrumentation, clinical criteria, patient demographics, etc.).

CLUSTER ANALYSES

Several cluster analyses were performed in an attempt to identify those QICs exhibiting "unusual" values for selected performance indicators. Clustering the QICs on the basis of the values of the performance indicators obtained from the data collected by each QIC can be used as a data quality monitoring procedure. The results of these analyses can be used to identify QICs for auditing.

CLUSTER ANALYSES USING LIFETIME NONUSE OF TOBACCO

SAS® EM™ was used to investigate similarities and differences in QICs' reported percentages of lifetime nonuse of tobacco. Prior analysis at WVMI showed that the overall percentage of the VA population of patients who report lifetime nonuse of tobacco is approximately 9%, and this percentage is more or less constant across VAMCs. Extreme departures of QIC-reported percentages from this overall percentage may be an indication of data quality problems or fraud on the part of QICs.

An initial cluster analysis was performed using SAS® EM's default parameter values and clustering methods. Wards method, using a cubic clustering criterion, was employed to generate the clusters. SAS® EM™ was allowed to automatically determine the number of clusters generated. The results included 12 clusters, resulting in an extremely high r-square value of 0.9999. The large number of clusters made interpretation difficult, and the high r-square is an indication of over-fitting of the data.

A second cluster analysis similar to the initial cluster analysis was performed, except that the number of clusters was user-specified as three. In essence, by controlling the number of clusters, a discriminant analysis was performed. The number of clusters (3) was chosen anticipating discrimination of QIC reported-percentages into three groups: one group of QICs exhibiting lower than expected percentages, another group of QICs whose percentages were nearly those expected, and a third group of QICs with higher than expected percentages. The resulting clustering produced an r-square value of 0.8133. However, the clusters generated by SAS® EM™ were not those anticipated.

The first cluster contained 60 QICs and exhibited a cluster mean of 7.5% for the percentage of lifetime nonuse of tobacco. This cluster is interpreted to contain those QICs whose percentages were slightly less than or near the expected overall percentage of 9%. The second cluster contained 13 QICs with a cluster mean of 21.5 %. This cluster is interpreted to contain those QICs whose percentages were more than twice the expected overall percentage of 9%. The third cluster contained 3 QICs with a cluster mean of 44.0%. The third cluster can be interpreted as grouping together those QICs whose reported percentage of lifetime nonuse of tobacco is extremely greater than the expected 9%. There exist at least four possible explanations for the anomalous percentages of the three QICs in cluster 3 – negligent or fabricated record abstraction, lax data collection, or anomalies due to sampling procedures and/or demographic distributions of smokers in the VA system. If the sampled records are randomly selected, it is to be expected that approximately 5% of the samples will result in samples that are not representative of the population of patients at a particular VAMC. Five percent of the 79 QICS examined is 3.95. We can expect that about four of the QICs should exhibit "unusual" values for their reported percentages. These results may warrant further investigation.

A third cluster analysis was then performed, similar to the previously discussed cluster analyses, except that the number of clusters was user-specified as two. The number of clusters, 2, was chosen anticipating discrimination of QIC reported-percentages into two groups: one group of QICs exhibiting percentages close to the expected percentages, and a second group of QICs with higher than expected percentages. The resulting clustering had an r-square value of 0.6764. The first cluster contained 69 QICs and exhibited a cluster mean of 8.9% for the percentage of lifetime nonuse of tobacco. The second cluster contained 7 QICs with a cluster mean of 35.6 %. The second cluster can be interpreted as grouping together those QICs whose reported percentage of lifetime nonuse of tobacco is extremely greater than the expected 9%. These results may warrant further investigation.

It appears that a cluster analysis limited to three clusters (actually a discriminant analysis with three groups) provides a reasonable fit to the data, and allows a reasonable interpretation of the QIC groupings. The results can also be used to select QICs for auditing.

CLUSTER ANALYSES USING SERVICE REFUSALS

SAS® EM™ was used to investigate similarities and differences in QICs' reported percentages of the indicator refusal (indicating whether a patient refused any tests or procedures recommended by a

physician). Prior analysis conducted at WVMI suggests that the overall percentage of the VA population of patients refusing any test or procedure is approximately 12%. This percentage is approximately constant across VAMCs, therefore, extreme departures of QIC-reported percentages from this overall percentage may be an indication of data quality problems or performance problems on the part of the QICs.

An initial cluster analyses was performed using SAS® EM's default parameter values and clustering methods. As in the analysis described above, Wards method, using a cubic clustering criterion, was employed to generate the clusters. SAS® EM™ was allowed to automatically determine the number of clusters generated. The results included 40 clusters, producing an extremely high r-square value of 0.9998. The large number of clusters made interpretation difficult, and the high r-square is an indication of overfitting of the data.

A second cluster analysis similar to the initial cluster analysis was performed, except that the number of clusters was user-specified to be three. The number of clusters, 3, was chosen anticipating discrimination of QIC reported-percentages into three groups - one group of QICs exhibiting lower than expected percentages, another group of QICs whose percentages were nearly those expected, and a third group of QICs with higher than expected percentages. The resulting clustering resulted in an r-square value of 0.7361. The first cluster contained 21 QICs and exhibited a cluster mean of 10.2% for the percentage of patients refusing at least one recommended test or procedure. This cluster is interpreted to contain those QICs whose percentages were slightly less than or near the expected overall percentage of 12%. The second cluster contained 51 QICs with a cluster mean of 19.2 %. This cluster is interpreted to contain those QICs whose percentages were near or somewhat greater than the expected overall percentage of 12%. The third cluster contained 7 QICs with a cluster mean of 32.47%.

A third cluster analysis was then performed, similar to the previously discussed cluster analyses, except that the number of clusters was user-specified as two. The number of clusters, 2, was chosen anticipating discrimination of QIC reported-percentages into two groups - one group of QICs exhibiting percentages close to expected percentages, and a second group of QICs with higher than expected percentages. The resulting clustering had an r-square value of 0.6377. The first cluster contained 51 QICs and exhibited a cluster mean of 13.9% for the percentage of patients who refused at least one recommended test or procedure. The second cluster contained 28 QICs with a cluster mean of 25.5 %. The second cluster can be interpreted as grouping together those QICs whose reported percentage of patients who refused at least one recommended test or procedure is somewhat greater than the expected 12%. The results of this analysis indicate that the use of only two groupings does not provide a high level of discriminatory power.

A fourth cluster analysis was subsequently performed, except that the number of clusters was user-specified to be five. The number of clusters, 5, was chosen anticipating discrimination of QIC reported-percentages into five groups - one group of QICs exhibiting lower than expected percentages, another group of QICs whose percentages were higher than those expected, and the other three groups containing QICs with percentages near the expected percentage of 12%. The resulting clustering resulted in an r-square value of 0.8878. The first cluster contained 32 QICs and exhibited a cluster mean of 12.8% for the percentage of patients refusing at least one recommended test or procedure. This cluster is interpreted to contain those QICs whose percentages were near the expected overall percentage of 12%. The second cluster contained 31 QICs with a cluster mean of 20.0 %. This cluster is interpreted to contain those QICs whose percentages were slightly greater than the expected overall percentage of 12%. The third cluster contained 10 QICs with a cluster mean of 27.1%. The third cluster can be interpreted as grouping together those QICs whose reported percentage of patients refusing at least one recommended test or procedure is somewhat greater than the expected 12%. The fourth cluster contained 3 QICs with a cluster mean of 4.8%. The fourth cluster can be interpreted as grouping together those QICs whose reported percentage of patients refusing at least one recommended test or procedure is less than the expected 12%. The fifth cluster also contained 3 QICs with a cluster mean of 36.7%. The fifth cluster can be interpreted as grouping together those QICs whose reported percentage of patients refusing at least one recommended test or procedure is extremely greater than the expected 12%.

For purposes of data monitoring and quality assurance, it appears that a cluster analysis using five clusters (actually a discriminant analysis with five groups) provides a reasonable fit to the data, allows a reasonable interpretation of the QIC groupings, and can be used to identify QICs for potential auditing.

MULTIVARIATE CLUSTER ANALYSES USING SIX PERFORMANCE INDICATORS

The cluster analyses reported in the previous two sections were univariate analyses. These types of analyses may be reasonable to assess performance of QICs on a single variable. To investigate wider performance issues (including fabriction of data) on the part of QICs, it would seem more reasonable to assess their performance, not by using a single indicator, but using a set of indicators. Thus a multivariate cluster analysis was performed. The QICs were clustered according to their reported percentages for six indicators. These include colon refusal rate (did patient refuse colon screening), fecal refusal rate (did patient refuse fecal testing), fluvac refusal rate (did patient refuse flu vaccination), mammgram refusal rate (did patient refuse mammogram), pap refusal rate (did patient refuse pap test), and pnumovac refusal rate (did patient refuse pneumonia vaccination).

An initial cluster analyses was performed using SAS® EM's default parameter values and clustering methods. QICs with missing values for any of the 6 indicators were excluded from the analysis (41 QICs were excluded). SAS® EM™ was allowed to automatically determine the number of clusters generated. The results included 38 clusters (each containing one QIC), resulting in a perfect r-square value of 1.000. The result that each cluster contained only one QIC and the high r-square is an indication of overfitting of the data. The results of this cluster analysis provide no useful information regarding the performance of the QICs.

A second multivariate cluster analysis was performed similar to the initial analysis, with the exception that QICs with missing values for the indicators were not excluded form the analysis. SAS® EM™ was allowed to determine the number of clusters to construct. The results contained seven clusters with an r-square of 0.6430. Clusters 4 and 6 differed in many regards from the other five clusters. Cluster 6 contained two QICs. Cluster 6 differed from the other 6 clusters in several measures. The mean for percentage of patients refusing a pap test was 25%, or about 3.5 times the highest refusal rate in any of the other clusters, and about 8 times higher than the mean refusal rate. The mean percentage of patients who were lifetime nonusers of tobacco was 47.8 %, which was twice the highest rate in any of the other 6 clusters, and about 5 times higher than the overall rate.

Cluster 4 contained seven QICs. Cluster 4 exhibited large differences from the other clusters on two measures. The refusal rate for pneumonia vaccinations was 18.9%, which was approximately three times the overall refusal rate and about three times the highest rate in any of the other clusters. The other six clusters exhibited rates not far from 6%. The colon refusal rate for cluster 6 was 0.8%, which is lower than the rates for the other six clusters. The colon refusal rate for cluster 6 was less than half of the lowest rate in the other six clusters, and about one-third of the overall rate.

The results of this analysis may indicate a need to audit two of the QICs. However, the fact that these two QICs did not differ greatly from the means of the other three indicators may indicate that the unusual values for pap refusal, mammogram refusal, and lifetime nonuse of tobacco may be specific to the VAMCs from which the records were abstracted. These results indicate the need for an analysis of rates by VAMC. Furthermore, an analysis of rates by VAMC and QIC

combinations may also be warranted.

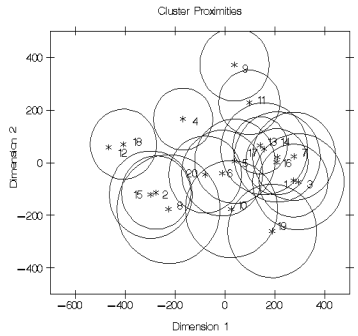
MODELING QIC EXPERIENCE AND IMPACT ON KEY PERFORMANCE INDICATORS

One of our interests is to develop a method for identifying items that may be difficult for QICs to abstract and/or agree as evidenced by low correlations in interrater reliability tests. Items that are difficult to abstract that impact KPI scores are of highest priority. Additionally, QICs with varying levels of experience may be more susceptible to abstracting difficulties—especially those in their first months of abstracting. If we can identify the items, QICs, and other contextualizing influences yielding less than optimal abstraction, then appropriate interventions can be implemented to improve performance.

To begin this model building WVMI first constructed a data set that contained the "key performance indicators" that were used to assess clinical performance in the FY 2001 EPRP reports. The data were then summated to reflect the review periods (i.e., each calendar quarter), VAMCs, and QICs conducting the abstractions. As mentioned above, it is important to our model that we measure the extent of each abstractor's experience; the summated dataset also contained information on the number of records abstracted in that review period and the QICs experience at the beginning of the review period (by number of months).

Each of the KPI variables were initially used as input data, along with the QIC specific data. Three initial cluster analyses were run using either an average linkage, Ward's, or centroid method and CCC to determine the number of clusters to retain. The centroid method yielded approximately 20 clusters. From the initial list of approximately 40 variables, about 25 had an importance level of 0. A second analysis was run which excluded these variables, with approximately the same number of clusters, but with smaller radiances.

Figure 3. QICs WITH OUTLIER KPI SCORES



In identifying these clusters, we are now in a position to assess the mitigating factors that may be impacting on the performance of QICs, especially those with minimal field experience. For example, we will compare each of these clusters against trended data performed by seasoned QICs. We will also consult additional data sources such as interrater training exercises and independent audits that are routinely performed on new QICs. Through a combination of these methods, we expect to build a model that will identify items needing special focus during training. The model will also be incorporated into our routine screening of data streams to help assure quality abstractions.

CONCLUSION

SAS® EM™ provides tools for exploring relationships in our data that point to better models of assessing data quality. The use of SAS® EM™ is supporting the creation of a more refined approach to model building: a method for identifying difficult medical record items and correlating performance with abstraction experience. Such a model yields an important quality improvement tool necessary for quick intervention through education and training. Although our primary objective for building and applying the models are for continuous quality improvement, we will also use the models for monitoring abstractors to help assure high quality performance.

REFERENCES

1. Forsythe, James H., Perlin, Jonathan B., Brehm, John, "Data Quality and Medical Record Abstraction in the Veterans Health Administration's External Peer Review Program." Proceedings of the Sixth International Conference on Information Quality, Elizabeth M. Pierce and Raissa Katz-Hass, eds., Massachusetts Institute of Technology, Cambridge, MA 2001.

ACKNOWLEDGMENTS

The authors express appreciation to Helen Cowie, M.S., WVMI, for her assistance in the preparation of this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

James H. Forsythe, Ph.D.
West Virginia Medical Institute
3001 Chesterfield Place
Charleston, WV 25304
Work Phone: 304-346-9864 x2265
Fax: 304-346-9863
Email: jforsythe@wvmi.org
Web: www.wvmi.org