A Matter of Classes: Stratifying Health Care Populations to Produce Better Estimates of Inpatient Costs

David B. Rein

Objective. To stratify traditional risk-adjustment models by health severity classes in a way that is empirically based, is accessible to policy makers, and improves predictions of inpatient costs.

Data Sources. Secondary data created from the administrative claims from all 829,356 children aged 21 years and under enrolled in Georgia Medicaid in 1999.

Study Design. A finite mixture model was used to assign child Medicaid patients to health severity classes. These class assignments were then used to stratify both portions of a traditional two-part risk-adjustment model predicting inpatient Medicaid expenditures. Traditional model results were compared with the stratified model using actuarial statistics.

Principal Findings. The finite mixture model identified four classes of children: a majority healthy class and three illness classes with increasing levels of severity. Stratifying the traditional two-part risk-adjustment model by health severity classes improved its R^2 from 0.17 to 0.25. The majority of additional predictive power resulted from stratifying the second part of the two-part model. Further, the preference for the stratified model was unaffected by months of patient enrollment time.

Conclusions. Stratifying health care populations based on measures of health severity is a powerful method to achieve more accurate cost predictions. Insurers who ignore the predictive advances of sample stratification in setting risk-adjusted premiums may create strong financial incentives for adverse selection. Finite mixture models provide an empirically based, replicable methodology for stratification that should be accessible to most health care financial managers.

Key Words. Risk adjustment, health care financing, inpatient costs, adverse selection

Traditional models of health care utilization, such as ordinary least squares (OLS) regression models and the two-part model, restrict the effect of predictive covariates to be equal for all patients in the data. This may be unrealistic as the effects of predictive variables such as gender, age, or diagnosis

may differ substantially between patients, particularly between groups of patients with different levels of illness. Sicker patients may be more difficult to care for than healthier patients even for routine medical care. For example, performing a routine physical for a child with autism may be more time consuming and therefore more costly than performing the same physical on a nondisabled child. However, traditional models rarely offer a methodological means to adjust the predictive covariates by patient characteristics.

This paper uses a finite mixture model to identify classes or bundles of patients that consume health care resources in a similar fashion. It builds from the simple premise that health care patients can be separated into different classes based on the severity of their health care needs and that the way in which patients consume health care differs based on these classes. Because patients are likely to differ in their use of services across classes, models such as this one that allow the effects of predictor variables to also differ across classes are likely to result in more accurate estimates. More accurate predictions of utilization can inform decisions regarding risk sharing within a number of reimbursement environments.

The idea that patient populations are comprised of subclasses with varying degrees of health and illness is not new either conceptually (McPherson et al. 1998) or methodologically (Deb and Trivedi 1997). However, conceptual work has struggled to find a quantitative tool to define classes, and methodological work has struggled to develop tools that are flexible and accessible enough to meet the demands of applied policy makers.

This paper demonstrates a straightforward application of a quantitative categorization system, finite mixture modeling, to define different severity classes in a health care population. It then uses these newly defined classes to improve the power of the classic two-part risk-adjustment model in predicting inpatient expenditures. Inpatient expenditures were chosen as the illustrative example because they are particularly costly, and the use of inpatient services has a large probabilistic component; only 13 percent of patients in this example utilize any care at all. However, the same model has been used to predict outpatient and prescription drug services using the same population data, resulting in the same conceptual results (Rein 2003).

Address correspondence to David B. Rein, Ph.D., RTI International (trade name of Research Triangle Institute), Division of Health Economics Research, 2951 Flowers Road South, Suite 119, Atlanta, GA 30341 or by e-mail at drein@rti.org.

DATA

The data for this study were drawn from a larger evaluation of Georgia's Child Medicaid program for all 829,356 Georgia children aged 21 years and under enrolled in 1999. Georgia Medicaid covers all inpatient services except some organ transplants in certain circumstances. Nonemergency, nonmaternity admissions require prior approval from the patient's primary care provider or referring specialist. In Georgia, Medicaid is run on a fee-for-service basis, meaning that medical providers submit a unique medical claim to the state for reimbursement of each service they perform. These claims contain a large amount of information, including patient diagnosis, and the cost of the service (MEDStat Group 2001). Individual claims can be matched, using anonymous unique identifiers, to person-level enrollment data containing demographic information about the patient. Because there are Medicaid programs in all 50 states, lessons learned using one state's Medicaid data may be widely applicable.

The data include only information from medical claims and thus do not include information from chart abstractions, records from long-term care, or over-the-counter medications. Children enrolled via the state child health insurance program (PeachCare for Kids) were excluded from the analysis because they differed from children enrolled in Medicaid in several important respects (Edwards, Bronstein, and Rein 2002). Adjustment claims used to void earlier records were matched to their original record, and each voided set was eliminated from the data. After cleaning, a random sample of 10 percent of the observations was selected and held aside for model validation purposes, creating a modeling sample of 746,036 and a validation sample of 83,320. The validation sample was not evaluated in any way until the conclusion of each modeling step.¹

METHODS

This analysis compares a traditional two-part cost-reimbursement model to an identically specified model with the exception that the alternative model is stratified by finite mixture classes estimated from a measure of patient health status. This section explains the traditional two-part model and its specification and the alternative model which uses finite mixture methodology to stratify this model by health status classes. It then explains how the predicted values from the alternative model were calculated, the evaluation statistics used to compare models, and the separate model specifications that were compared.

Two-Part Model

The traditional two-part model estimates the consumption of health services using separate equations for the decision to seek care, made by the patient, and the decision about how much care is used, made by the physician (the two components of Zweifel's classic principal-agent theory; Zweifel 1981). The two-part model has been the standard of risk adjustment methodology since Manning, Duan, and Rogers (1987) determined it to be better than several competing model formations.

The first part of the model estimates the probability of the patient (the principal) to seek care using a standard logistic regression model. The second part of the model estimates the amount of care received for only those patients who received care, a decision made by the provider (the principal's agent). The second part is traditionally estimated using a continuous log-linear OLS regression model (Manning and Mullahy 1999), although more recently other distributional assumptions have been used to good effect (Blough and Ramsey 2000).² Because this paper focuses on issues of sample stratification, the more accessible log-linear model is used.

The models used in this analysis estimated both the logistic model of the probability of seeking care and the log-linear model of the expected dollar-denominated value of care as a function of exogenous variables drawn from Georgia Medicaid claims data that measured the components of the Anderson-Aday Behavioral Model of Health Utilization (Aday 1998). Specifically, each dependent variable was estimated as a function of dummy variables measuring age; race, African American as compared with other races; ethnicity, Latino as compared with other ethnicities; female gender and female gender of reproductive age; months of 1999 enrollment; urbanization; Medicaid enrollment status; 21 diagnostic indicators of disease used in Georgia to identify needy children; and an indicator of having more than one special need diagnosis (age 19–21 years, white race, non-Latino ethnicity, male gender, full-year enrollment, most urban area, eligibility via income, and no diagnosed special need diagnoses were used as the reference cases).

Following standard methodology (Manning and Mullahy 1999), Duan smearing was then used to adjust for retransformation bias, which occurs when log-normalized predicted values are exponentiated back to the overly dispersed, nonnormal distributional dollar scale of health expenditures (Duan et al. 1983; Diehr et al. 1999). Model estimates of individual predicted probabilities of service and expected value of costs were then multiplied to arrive at individual estimates of expected costs.

Alternative Model: Stratifying the Two-Part Model by Finite Mixture Classifications

The alternative model stratified each portion of the traditional two-part model by health severity classes. A finite mixture model assigns individual values of an observed variable to different unobserved subclasses. The assumption of the model is that these different classes were mixed together to create the data now observed. By modeling the distribution of the observed variable in the right way, probability estimates of individual membership in each unobserved class can be obtained. As shown below, the value of y for any single individual (i) is a function of that individual's probability of class membership (π) in c number of distinct subpopulation distributions multiplied by the appropriate function of y corresponding to each class:

$$f(y_i) = \sum_{j=1}^c \pi_i f_i(y_i) \quad (1)$$

In the absence of explanatory variables, an individual's value of y is simply equal to the sum of the mean of y in each subpopulation class multiplied by the individual's probability of membership in that class. This research uses a finite mixture model to identify different clusters of patients' health severity scores, with each cluster exhibiting its own estimated mean and standard deviation. The model provides empirical evidence regarding the number of subpopulation classes needed to best explain the distribution of y .

The mathematical estimation of the number of subclasses is quite complex and is treated thoroughly and in great depth in other sources (Muthen and Muthen 1998; McLachlan and Peel 2000). Conceptually, in the absence of strong theory, the number of classes is estimated using model diagnostic statistics such as the Bayesian information criterion (BIC) to compare models with different numbers of subcomponent classes. The BIC (smaller values are preferred) provides a measure of overall model fit based on improvements in the likelihood function while rewarding the researcher for parsimony. The model with the number of classes that results in the smallest BIC statistic is the model that is most likely to have generated the observed data. From a theoretical perspective, the finite mixture model used in this paper measured an unobserved categorical variable describing health severity class. Mechanically, the model identified empirically distinguishable data clusters in the distribution of Chronic Disability Payment System (CDPS) scores which were then labeled by the researcher and put to use to stratify utilization models.

As a stratification technique, finite mixture modeling has two main advantages. First, mixture modeling provides a quantifiable metric, the BIC statistic, to determine optimal categorization. Second, finite mixture models easily allow the assignment of partial class memberships to members that might not exhibit obvious signs of membership in any single class.

In this research, models with one, two, three, four, five, and six classes of health users were estimated or attempted and then compared. Because the likelihood function may be sensitive to changes in starting values, models were tested multiple times with different starting values each time. To test whether the same classes identified in the modeled data could also be identified in an unfitted sample, models with one through six classes were also tested in the validation sample. This was done by restricting parameter estimates (class health severity score means and variances) in the validation data models to be equal to those found in the fitted data and comparing models with one through six classes using the Akaike information criteria (AIC) statistic (similar to the BIC but preferred for out-of-sample validation [Deb and Trivedi 2002]). The model with the smallest BIC statistic that was insensitive to changes in starting values and validated using the holdout data was used to stratify each portion of the two-part model.

The CDPS score, the health status variable from which subclasses were determined, is a measure of the cumulative amount of patient illness over a given time period. It assigns weights to diagnoses and demographic variables based on those variables' propensity to generate costs in other state Medicaid databases (Kronick et al. 2000). Medicaid patients with a CDPS score of 1 are expected to incur the average overall Medicaid costs; patients with a score of 2 are expected to incur costs that are twice the average; and so on. The CDPS was selected over other systems because of its specific calibration to a Medicaid population and its public domain availability. The CDPS has been used to predict medical expenditures by state Medicaid programs in Pennsylvania, New Jersey, Tennessee, Virginia, Oregon, Colorado, Michigan, and Utah, although its success has yet to be fully evaluated. Minnesota, Arizona, Ohio, Oklahoma, and Orange County California are currently exploring using the scoring system (for more information visit the CDPS website: <http://medicine.ucsd.edu/fpm/cdps/>).

The finite mixture model was based on four quarterly CDPS measures, as opposed to one annual measure, in order to distinguish acute episodic illness from chronic disease. When patients experienced less than a full-year's enrollment, quarterly CDPS scores were imputed using a last observation carried forward (LOCF) methodology.

Calculating Predicted Values of Health Expenditures

To compare the traditional model with its stratified alternative, two predicted values were estimated for each individual, one corresponding to the standard two-part model, and an alternative value estimated for the stratified model. To achieve alternative model values, different versions of the two-part model were estimated for each of the health severity classes. Individual probabilities of membership in each health service class were multiplied by individual expected expenditure estimates for that class. These values were summed to create individual expected expenditure estimates. For example, the i th child's probability of any inpatient service was derived by multiplying her probability of belonging to class 1 by her probability of inpatient service predicted by the logistic model for class 1 members, and then adding that product to her probability of belonging to class 2 multiplied by her probability of inpatient service derived from the logistic model for class 2 and so on, so that

$$IP_i = (CP1_i IP1_i) + (CP2_i IP2_i) + \dots + (CPn_i IPn_i) \quad (2)$$

where IP equals inpatient service probability, $CP1_i$ – CPn_i equals probabilities of membership in class 1 through class n , and $IP1_i$ – IPn_i equals inpatient service probability predicted by the models for classes 1, ..., n .

Expected values of the log of inpatient expenditures for each individual were created similarly. Each class's expected value was multiplied by the probability of belonging to that class, following the same logic used to find the predicted inpatient probability. Next, the log-scale expected values were re-transformed into dollar values, adjusting for retransformation bias using Duan smearing adjustment for heteroscedasticity (Duan et al. 1983). Finally, individual dollar scale expected values were multiplied by individual predicted probabilities of service to yield individual-level predicted values of inpatient expenditures. These alternative estimates from the stratified model were compared with those generated by the traditional two-part model.

Evaluation Statistics

Three actuarial statistics were used to compare the final predicted expenditure value from the traditional and alternative models. First, the predictive ratio, equal to the sum of the predicted values divided by the sum of the actual values, was used to determine how well each model predicted cumulative expenditures over all individuals. The predictive ratio ranges from zero to infinity with values closer to 1 preferred. Values less than 1 indicate that the model systematically under-predicts expenditures, while values greater than 1

indicate over-prediction (Cumming et al. 2002). Second, the mean absolute prediction error statistic, measuring the average absolute deviance between the predicted and actual values, was used to measure the degree to which individual predictions varied from their actual values. Third, the R^2 statistic, ranging from zero to 1 with higher values preferred, was used to measure the percentage of total variance of inpatient expenditures explained by each model.

Model Comparisons

Several models were compared in this analysis. First, the traditional two-part model was compared with an alternative in which each part of the model was stratified by the finite mixture classes. Second, to determine the source of model improvement, the two-part model was compared with a model in which only the first part was stratified and to a model in which only the second part was stratified by finite mixture classes. Third, the second part of the two-part model (the expected dollar value of services for those who use service) was compared with its alternative stratified counterpart, because these models have the greatest applicability in retrospective cost-reimbursement scenarios where service has already occurred and thus predicting the probability of service is not important.

Many Medicaid patients experience only a partial year's enrollment (Table 1). If temporary enrollment influences the ability to accurately estimate a patient's health class, then the alternative model may not be applicable in situations where a great deal of temporary, or episodic, enrollment occurs, decreasing its generalizability. To test whether length of enrollment influenced the preference between the traditional and the alternative model, the models were compared using restricted samples of children with one-quarter, two-quarters, three-quarters, more than three-quarters but less than a full-year's enrollment, and a full-year's enrollment. For all comparative statistics, only results from the validation sample were considered.

Some disagreement exists regarding the additional value (if any) of the two-part model with Duan smearing over a single equation OLS regression of a dollar scale variable. Whichever method is preferred statistically; in practice this simpler OLS approach is likely more commonly applied. For comparison purposes, a single equation OLS model was estimated against the entire sample of observations (those with and without health expenditures) with the same covariates as the second part of the two-part model. Because this model was inferior to the two-part model on all diagnostic criteria its results were omitted.

Table 1: Summary Statistics of Patient Characteristics

<i>Variable or Group of Variables</i>	<i>Mean or Percentage</i>
Dependent variables	
Percentage with any inpatient expenditures	13
Mean inpatient expenditures over all children	502
Mean inpatient expenditures for children with inpatient expenditures	4,081
Independent variables	
Age (years)	
0	15
1–5	30
6–10	22
11–15	16
16–21	16
Quarters of enrollment	
1 or less	14
1–2	15
2–3	16
3–4 (but not full year)	12
Full-year's enrollment	43
Black	50
Latino	4
White	31
Other and unspecified	15
Female	53
Female over 12 years old	18 (of all), 34 (of females)
Urban group	
Large city	62
Small city	13
Rural	25
Enrollment eligibility	
Income qualified	93
Supplemental security insurance	5
Foster care	2
Diagnosis	
One of the 21 special need diagnoses	15
Two or more of the 21 diagnoses	3

RESULTS

Finite mixture models with one through six classes of health users were attempted. For models with four or fewer classes, the model's BIC statistic was recorded and compared. Models with more than four classes failed to converge because of a computational inability to invert the Fisher information

matrix. Unpublished research indicates that convergence failure may be related to achievement of the maximum number of classes in a given population (Muthen 2002). Of the four models that converged, the model with four health severity classes produced the lowest (preferred) BIC statistic. This result was replicated in the validation sample (Table 2). This result suggests a model with four classes: a majority class of healthy patients and three distinct classes of patients exhibiting increasing levels of illness (descriptive results available by request). These classes map to immediately recognizable differences in inpatient service consumption. Greater proportions of patients in higher severity classes used inpatient services, and the mean cost of these services increased as severity class increased (Table 3). Virtually identical results were observed in the validation sample.

Selection statistics from both the logistic and the log-linear portion of the two-part model favored the alternative, stratified formation over the traditional form. The alternative model achieved more precise point estimates of individual probabilities of service and expected values of service.

Overall fit statistics indicated that the alternative stratified model resulted in better predictions of health expenditures than the traditional un-

Table 2: BIC (Fitted Data) and AIC (Validation Data) Statistics for Models with 1–5 Groups

<i>Number of Latent Classes</i>	<i>BIC Statistic (Fitted Sample)</i>	<i>AIC Statistic (Validation Sample)</i>
1	6,442,056	359,635
2	– 415,111	– 19,769
3	– 474,533	– 31,601
4	– 588,733	– 64,151
5	Could not be estimated	Could not be estimated

Table 3: Percent Categorized in Each Group, Percent Using Inpatient Services and Inpatient Expenditure Distribution Characteristics from Each Class

	<i>Class 1</i>	<i>Class 2</i>	<i>Class 3</i>	<i>Class 4</i>
Percent in class	63.4	26.0	9.0	1.6
Inpatient percent	7	21	23	44
Mean	144	688	1,457	6,416
SD	869	1,339	1,885	2,647
Min	0	0	0	0
Max	256,577	316,143	206,208	535,225

stratified form over all observations and in each health severity class in the validation data (Table 4). The alternative model produced a better predictive ratio, lower error, and greater explanatory power over all observations. The traditional model overpredicted total expenditures for class 1 (healthy children) by almost a factor of 2 and underpredicted total inpatient expenditures for the sickest by a much larger degree than the alternative model. Although the alternative model produced a slightly higher MAPE statistic in classes 3 and 4 than the traditional model, the magnitude of the difference was small compared with the large improvement in the predictive ratio and R^2 statistics of the alternative model.

We also considered the degree of improvement which resulted from stratifying each part of the two-part model by finite mixture classes. Stratifying only the first part of the model resulted in no change in the predictive ratio, an 8 percent decrease in the MAPE statistic, and an improvement in the R^2 statistic from 0.17 to 0.19. Stratifying only the second part of the model also resulted in no change in the predictive ratio, a smaller (5 percent) decrease in the MAPE statistic, and a large improvement in the R^2 statistic from 0.17 to 0.24. Restricting the sample to only those who used service and applying stratification to the expenditure portion of the two-part model (as might be done in retrospective cost reimbursement) resulted in an improved predictive ratio and MAPE statistic and a strongly improved R^2 compared with a similarly restricted traditional model. To test the impact of missing quarters of enrollment data the model's performance was also compared between groups

Table 4: Comparison of Predictive Ratio, MAPE, and R^2 Statistics from the Traditional and Alternative Models over All Patients and in Each Health Severity Group

<i>Class</i>	<i>Model</i>	<i>Predictive Ratio</i>	<i>Mean Absolute Predicted Error</i>	<i>R²</i>
<i>Results from validation data</i>				
All	Traditional	0.96	627	0.17
	Alternative	1.02	567	0.25
Class 1	Traditional	1.94	278	0.19
	Alternative	1.23	201	0.20
Class 2	Traditional	0.98	852	0.12
	Alternative	1.03	784	0.15
Class 3	Traditional	0.66	1,585	0.14
	Alternative	0.88	1,634	0.19
Class 4	Traditional	0.37	4,965	0.18
	Alternative	1.01	5,268	0.27

of patients with different lengths on enrollment; patients with one, two, three, greater than three but less than a full year, and four full quarters of enrollment. Again, the alternative model yielded better results than the traditional model for each subset of patients (tables available from the author by request).

LIMITATIONS

The model presented here, like all models, suffers from several limitations. First, because of limited data availability, the model was estimated only for children in Georgia Medicaid in 1999. This population differs from other populations in many ways, including their age, income, racial mix, illness severity, reimbursement system, and most importantly their heterogeneity of need. Further, utilization in a fee-for-service system such as the one used by Georgia Medicaid is likely to differ from utilization in a managed care system. However, finite mixture models have also resulted in improved predictions of utilization in other data (Deb and Trivedi 1997, 2002; Deb and Holmes 2000). Further, the model improvements found here result from the recognition of population heterogeneity. Although class estimates and regression coefficients would almost certainly differ in different populations, the strategy of stratification based on mixture model defined severity classes should be effective wherever heterogeneous classes of patients are pooled.

Second, LOCF was used to impute missing values of data over other alternatives such as multiple imputation. LOCF assumes that a patient's missing values follow the same distribution as their observed values. This assumption is supported by the high correlation ($\rho = .76$, $p < .01$) between quarterly CDPS scores in patients with multiple quarters of data. The main advantage of this method is its simplicity and computational ease. The main disadvantages are that first it assumes missing values occur at random when it is likely they do not, and second by increasing sample size, LOCF reduces the standard error of the estimates (Wood et al. 2005). In the case of this research deflated standard errors could potentially bias the mixture model toward a greater number of health severity classes. However, the high correlation across waves, and the fact that the model yields substantial predictive improvement for each group of patients when the model is subdivided by quarters of enrollment suggests that this bias is modest and that the effect on the results of the analyses is small. Future research should investigate the use of alternative methods such as multiple imputation for the replacement of missing values, particularly when modeling data with fewer observation.

Third, this study was applied to all individuals within a given age range enrolled in Medicaid over a given time. Future studies may want to drill down and concentrate more fully on particular types of patients. For example, pregnancies could be eliminated from the analysis, or the analysis could be limited to those with a previous year's utilization information or to those who have previously been diagnosed with a chronic health condition. Preselecting the sample may lead to classes with interpretations different from the ones found here, classes that may be more useful for policy planning or rate setting.

Fourth, patient self-perception of evaluated need is an important predictor of health care utilization (Hornbrook and Goodman 1996) that has been excluded from this model. The purpose of this exercise was to craft an alternative model using only administrative claims data that could be replicated by other localities without the need for additional data collection. However, the addition of survey-based measures of patient characteristics (or any type of more refined data) is likely to improve any risk-adjustment model.

DISCUSSION

This paper presented a new approach to cost-reimbursement modeling based on stratifying the traditional two-part cost-reimbursement model by finite mixture classes defined using a measure of patient health severity. The approach was predicated on the hypothesis that health care users are drawn from different subclasses based on health severity, that members of distinct groups consume health care in different ways, and that these differences matter in the prediction of both the probability and the expected value of care. The paper presented a technique to measure these classes and use them within the framework of the traditional two-part model. Doing this improved the model's overall R^2 from 0.17 to 0.25, a substantial improvement given the very high variance in health expenditures. This is not surprising because the finite mixture model adds additional information to the model. Still, stratifying by mixture classes improves the predictive power of the model substantially more than simply adding the CDPS score to the model as an explanatory variable (results available upon request).

Although some degree of predictive improvement was gained by stratifying the first part of the model, the majority of the model improvement came from the application of stratification techniques to the second (expenditure) portion of the model. Further, stratifying the model by finite mixture classes resulted in model improvement when looking only at the second part of the

model, which predicted the costs of services for those who used them. This is important because the second part of the model drives retrospective cost reimbursement, the most likely application of this work.

Restricting the data to patients with different enrollment periods did not change the main finding of the research, that stratification by finite mixture classes improves cost-reimbursement model predictive power, as model stratification led to improvements in predictive power regardless of the number of quarters a patient was enrolled. Interestingly, applied to this data the model was a better predictor of expenses for those with a partial year's enrollment than it was for those with a full-year's enrollment. This is perhaps explained in part by the higher probability of Georgia Medicaid recipients with a partial year's enrollment to use inpatient services (because an inpatient episode can trigger Medicaid eligibility). Still, this unexplained finding requires further examination before generalizing to other data.

In traditional markets, prices are set through the many decentralized buying and selling decisions of consumers and providers, using price competition to match consumer preferences with producers' capacity and willingness to supply goods and services. However, as has long been noted, the presence of health insurance distorts the normal functioning of the price mechanism as a regulator of health care supply and demand (Arrow 1963). In the absence of price competition, reimbursements must be set centrally by an organizing payer, such as a private insurance company or the state, creating significant challenges. Systems that set reimbursements that account for the diversity in patient presentation of illness and provider practice patterns (Gastonis et al. 1993) can be extremely helpful. In the absence of such a system, some degree of adverse selection is likely inevitable.

Forecasting future health care costs is extremely difficult, and there is wide consensus that current risk-adjustment models do not handle this problem well (Newhouse 1998). Given the difficulties of predicting future consumption, models such as this one should be used concurrently or retrospectively to evaluate elements of risk in insured populations with the intent of informing risk sharing arrangements with networks that care for the most difficult patients (Shewry et al. 1996) or designing retrospective two-part pricing adjustments (Newhouse 1998).

The main strategy employed by this paper is simply a form of sample stratification. Finite mixture modeling was chosen to determine stratification classes because of its empirical reliance on maximum likelihood estimation using test statistics to arrive at class memberships. Doing so resulted in substantial improvement power of the traditionally used cost-reimbursement

model. However, practitioners who may be uncomfortable with mixture methodology should not miss the central lesson of this work: patients are heterogeneous in their health care needs and presentation. Successful financing and care provision strategies will acknowledge these differences and will use sample stratification techniques to apply them in cost-reimbursement policy.

ACKNOWLEDGMENTS

Special thanks to E. Michael Foster and Gregory B. Lewis for their encouragement, review, and support. Funding for this research was supplied by RTI International, the Georgia Health Policy Center of Georgia State University, and through a dissertation grant supplied by the Agency for Healthcare Research and Quality (AHRQ R03 HS13286).

NOTES

1. All data were gathered and analyzed prior to April 14, 2003, and therefore were not subject to the HIPAA privacy rule. Still, data use agreements were obtained from the Georgia Department of Medical Assistance, which permitted the use of this data for research purposes, and would qualify the data used in this research as a limited dataset under current HIPAA guidelines.
2. The primary advantages of the traditional method are its computational ease, its interpretability, and its avoidance of misspecification problems related to assuming the wrong data distribution. Maximum likelihood and/or quasi-likelihood estimation requires the assumption or estimation of the data distribution of the dependent variable, and model results may suffer to the extent that this distributional assumption or estimation is incorrect (Manning and Mullahy 1999).

REFERENCE

- Aday, L. A. 1998. *Indicators and Predictors of Health Services Utilization*. 2d edition. Chicago: Health Administration Press.
- Arrow, K. J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *The American Economic Review* 53 (5): 941–73.
- Blough, D. K., and S. D. Ramsey. 2000. "Using Generalized Linear Models to Access Medical Care Costs." *Health Services and Outcomes Research Methodology* 1 (2): 185–202.
- Cumming, R. B., D. Knutson, B. A. Cameron, and B. Derrick. 2002. *A Comparative Analysis of Claims Based Risk Assessment for Commercial Populations*. Minneapolis, MN: Society of Actuaries.

- Deb, P., and A. M. Holmes. 2000. "Estimates of Use and Costs of Behavioural Health Care: A Comparison of Standard and Finite Mixture Models." *Health Economics* 9: 475-89.
- Deb, P., and P. K. Trivedi. 1997. "Demand for Medical Care by the Elderly: A Finite Mixture Approach." *Journal of Applied Econometrics* 12: 313-36.
- . 2002. "The Structure and Demand for Health Care: Latent Class versus Two-Part Models." *Journal of Health Economics* 21 (4): 601-25.
- Diehr, P., D. Yanez, A. Ash, M. Hornbrook, and D. Y. Lin. 1999. "Methods for Analyzing Health Care Utilization and Costs." *Annual Review of Public Health* 20: 122-44.
- Duan, N., W. Manning, C. Morris, and J. Newhouse. 1983. "A Comparison of Alternative Models for the Demand of Medical Care." *Journal of Business Economic Statistics* 1: 115-26.
- Edwards, J. N., J. Bronstein, and D. B. Rein. 2002. "Do Enrollees in 'Look-Alike' Medicaid and SCHIP Programs Really Look Alike?" *Health Affairs* 21 (3): 240-8.
- Gastonis, C., S. L. Normand, C. Liu, and C. Morris. 1993. "Geographical Variation of Procedure Utilization: A Hierarchical Model Approach." *Medical Care* 31 (5, suppl): YS54-9.
- Hornbrook, M. C., and M. J. Goodman. 1996. "Chronic Disease, Functional Health Status, and Demographics: A Multi-Dimensional Approach to Risk Adjustment." *Health Services Research* 31 (3): 283-307.
- Kronick, R., T. Gilmer, T. Dreyfus, and L. Lee. 2000. "Improving Health Based Payment for Medicaid Beneficiaries: CDPS." *Health Care Financing Review* 21 (3): 29-64.
- Manning, W. G., N. Duan, and W. H. Rogers. 1987. "Monte Carlo Evidence on the Choice between Sample Selection and Two-Part Models." *Journal of Econometrics* 35 (1): 59-82.
- Manning, W. G., and J. Mullahy. 1999. "Estimating Log Models: To Transform or Not to Transform?" *Journal of Health Economics* 20 (4): 461-94.
- McLachlan, G., and D. Peel. 2000. *Finite Mixture Models*. New York: John Wiley & Sons Inc.
- McPherson, M., P. Arango, H. Fox, C. Lauver, M. McManus, P. W. Newacheck, J. M. Perrin, J. P. Shonkoff, and B. Strickland. 1998. "A New Definition of Children with Special Health Care Needs." *Pediatrics* 102 (1, part 1): 137-40.
- MEDSTAT Group Inc. 2001. *Georgia Division of Medical Assistance, System 2 Field Descriptions*. Ann Arbor, MI: The MEDSTAT Group Inc.
- Muthen, L. K. 2002. *Mplus Discussion Group*. Los Angeles: Muthen & Muthen Inc. Available at: <http://www.statmodel.com/discussion/messages/13/51.html?1014659534>
- Muthen, L. K., and B. O. Muthen. 1998. *Mplus Statistical Analysis with Latent Variables: User's Guide*. Los Angeles: Muthen & Muthen.
- Newhouse, J. P. 1998. "Risk Adjustment: Where Are We Now?" *Inquiry* 35: 122-31.
- Rein, D. B. 2003. *Modeling the Health Care Utilization of Children in Medicaid*. Atlanta: Proquest Microfilm.

- Shewry, S., S. Hunt, J. Ramey, and J. Bertko. 1996. "Risk Adjustment: The Missing Piece of Market Competition." *Health Affairs* 15 (1): 171–81.
- Wood, A. M., I. R. White, M. Hillsdon, and J. Carpenter. 2005. "Comparison of Imputation and Modeling Methods in the Analysis of a Physical Activity Trial with Missing Outcomes." *International Journal of Epidemiology* 34 (1): 89–99.
- Zweifel, P. 1981. "Supplier-Induced Demand in a Model of Physician Behavior." In *Health, Economics, and Health Economics*, edited by J. van der Gaag, and M. Perlman, pp 245–67. Amsterdam: North-Holland.