A Method for Constructing Case-Mix Indexes, with Application to Hospital Length of Stay

Richard H. Shachtman, Steven M. Snapinn, Dana Quade, Deborah A. Freund, and Alan K. Kronhaus

This article presents the methodological development of an index for case-mix adjustment of hospital data exemplified by our construction of an index for studying length of stay. We describe the development and evaluation of this index, including internal and external validation procedures, and show an example of its use in a policy-relevant context by applying it to the analysis of length-of-stay differences between investor-owned and voluntary hospitals. Some advantages of this approach to adjusting for case mix are (1) applicability to many hospital or patient output measurements/diagnostic scheme situations; (2) usefulness in reducing heterogeneity in other case-mix adjustments, e.g., the Diagnosis-Related Group (DRG) approach; (3) interpretation possibilities; (4) production of a single score for each patient/hospital; (5) statistical approach allowing more accurate and reliable interpretation of hospital and patient output measurements, (6) ability to deal with hospital deaths; and (7) consideration of the complete set of secondary diagnoses. We also suggest other possible uses of this approach.

INTRODUCTION AND LITERATURE REVIEW

Policy decisions which necessitate comparing hospital output measurements, such as length of stay (LOS), require consideration of differ-

---

This research was supported by National Center for Health Services Research Grant #5 R01 HS04175.

Address correspondence and requests for reprints to Richard H. Shachtman, Ph.D., Professor and Director, The LOH Project, Department of Biostatistics, School of Public Health, University of North Carolina, Rosenau Hall 201H, Chapel Hill, NC 27514. Steven M. Snapinn, Ph.D. is a biometrician at Merck, Sharp & Dohme Research Laboratories; Dana Quade is a Professor in the Department of Biostatistics at The University of North Carolina; Deborah A. Freund, Ph.D. is an Assistant Professor, Department of Health Policy and Administration, University of North Carolina; and Alan K. Kronhaus, M.D. is President of the KRON Medical Corporation.

ences in case mix. As an illustration, administrators and clinicians cannot institute effective and equitable programs to reduce LOS, and policymakers cannot monitor the performance of such programs adequately, without first making appropriate adjustments for case mix.

Several case-mix adjustment schemes, with a variety of different objectives, have appeared in the literature. All of these designs have attempted to classify patients into homogeneous groups, often by utilizing some statistical technique to reduce the multidimensional set of measurements on each patient in a hospital in order to classify him/her into a particular category. In contrast, we have developed a new general method which produces a case-mix index for each hospital. Depending on the objectives, this indexing approach may be used independent of, in place of, or in addition to, other schemes.

In this article, we present the methodological development of our indexing approach and illustrate it by producing an index, based on a national probability sample, for use in considering policy questions about LOS. We describe its construction and validation, and refer to an application examining LOS differences between investor-owned and voluntary (not-for-profit) hospitals. We stress, however, that although we have worked specifically with LOS, and with several specific databases, our method can readily be adapted to adjust for case mix when analyzing cost or any other patient outcome; and it can be used with any database, or any patient or diagnosis coding system.

It is convenient to think of the factors which contribute to the observed variation in LOS as belonging to one of two mutually exclusive groups: those which are demographic or contribute to the disease status or degree of illness of the patient (patient-level factors), and those which represent the hospital environment or determine how a patient of given status is cared for from hospital admission to discharge (hospital-level factors).

The underlying strategy of our index method is to predict each individual patient's LOS on the basis of patient-level factors and then to subtract the resulting predicted value from the actually observed LOS. If this difference, or residual, is positive, the patient has had a longer stay than expected; if negative, the stay was shorter. Random variation might explain part of any such discrepancy in an individual patient. However, if the average residual (AR) for all patients in a hospital is significantly positive, this presumably reflects hospital-level factors which tend to prolong LOS; similarly, a negative AR suggests the presence of factors which shorten LOS.

We now describe briefly some of the existing schemes for case-mix adjustment, with the aim of highlighting how their attributes differ

from those of the index we have developed. A comprehensive review is available elsewhere [1].

The system of DRGs (Diagnosis-Related Groups) [2,3] designed to produce groups of patients homogeneous with respect to some measure of resource use (LOS or cost), and recently implemented as part of the Medicare prospective payment process, is undoubtedly the best-known method of patient classification. Two sets of DRGs have been developed, each of which begins by grouping patients into its own set of MDCs (Major Diagnostic Categories). In the older version, based on ICDA-8 (International Classification of Diseases—Adapted) codes [4], the DRGs are based on 83 MDCs organized by etiology; in the newer version adopted for Medicare using ICD-9-CM (International Classification of Diseases—Clinical Modification) codes [5], the DRGs are based on 23 MDCs organized by organ system. Within each MDC, an algorithm groups patients into DRGs, determining them in part by the statistical criterion of minimizing the within-DRG variance. The system deems patients within a DRG to be similar with respect to the measure of resource use, either length of stay (ICDA-8) or billing charges (ICD-9-CM). The DRG system takes into account severity of illness in a limited way by including indicator variables for the presence and absence of surgery and (in the ICD-9-CM version only) some comorbidity.

It is important to incorporate measures of severity into an index or classification system, since differences in severity may influence LOS. Most classification systems such as the DRGs, however, lack a comprehensive measure of the severity of the patients' illnesses. To remedy this defect, the Patient Severity Index (PSI) [6] scores a patient's severity (1 = least severe to 4 = death) on seven criteria, based on medical record data, which are implicitly averaged by the rater. Although this method is subjective, since only clinical judgment is used for scoring each patient and poor patient care can influence the score, interrater reliability is high.

Disease Staging [7] is another classification system which takes into account severity of illness. Its computerized staging algorithm, which grades patients' severity on a scale of 1 to 4 using clinical criteria based on available discharge abstracts, is more objective than the PSI. However, discharge abstracts are much less rich in detail than the commonly available medical records.

Finally, Patient Management Categories (PMCs) [8]—the title indicates their different objective—capture standard medical practice more effectively than the other methods described here, but it is not clear how easily they can be constructed using readily available archi-

val data. The computer algorithm designed to classify patients in 750 PMCs is not yet available for testing. PMCs group patients by reason for admission, specify what is required (independent of existing practice patterns) to treat these patients, and then determine the cost of this care. Actual cost data are allocated to PMCs on the basis of cost in a sample of Western Pennsylvania hospitals.

INDEX FOR CASE-MIX ADJUSTMENT

The method which we have developed is an index (rather than a patient classification system) which adjusts for case mix in a way that has advantages over other indexing or classification methods. We now detail some of these advantages. Our technique for constructing a case-mix index (1) is applicable to many patient and hospital output measurements (dependent variables) and many diagnostic coding schemes, regardless of the database used; (2) is useful in reducing heterogeneity in other case-mix adjustment approaches (e.g., the DRG scheme), and thus is employable in place of, or in addition to, such schemes; (3) allows interpretation possibilities using a reverse transformation to state results in the original units of measurement, e.g., days for LOS; (4) has the capacity to produce a single score for each patient and, *a fortiori*, each hospital, thus reducing the ubiquitous multidimensional set of patient measurements; (5) uses a transformation and parsimonious regression to normalize residuals for statistical testing, thus allowing more accurate and reliable interpretation of the relation between hospital and patient output measures and independent variables; (6) deals with deaths in the hospital; and (7) considers the complete set of secondary diagnoses. The policy importance of item (5) becomes increasingly clear when one recognizes that modest errors in measuring output variables for groups of patients, e.g., DRGs, could result in large differences in reimbursement amounts or interpretation of LOS differences on a national database. In particular, the index also enables us to answer questions of the following type: given the same disease status, how much longer would a patient stay in hospital B than in hospital A? We have not found an index designed for this purpose in the literature.

Like the DRGs, our method uses readily available data. However, it improves upon the DRG scheme by incorporating a more comprehensive adjustment for severity of illness. Although our database, in common with most others, does not have all of the information needed for the PSI, we account for severity of illness by using the complete list

of secondary diagnoses which provides a complete coding of comorbidity, and for surgical patients, the duration of anesthesia. We utilized the same MDCs as in the ICDA-8 version of the DRGs for the starting point of our case-mix index, but then diverged from the DRG development in several ways (discussed below). We devised a new statistical transformation that improves normality and thus allows for more legitimate hypothesis testing. (The Medicare index derived from DRGs uses a log transformation; our transformation follows a normal distribution more closely [9].) This transformation makes our model better able than the DRG system to account for in-hospital deaths. Finally, instead of adjusting for case mix by means of analysis of covariance, we use analyses of residuals. This provides conceptual clarity in distinguishing the independent variables of interest (hospital-level factors) from the covariables (patient-level factors), and helps avoid the ecological fallacy.[1]

DATA

The data used to develop our original index were collected by the Study on the Efficacy of Nosocomial Infection Control (SENIC) [10]. The SENIC data constitute a national probability sample of 170,000 patients admitted to 338 short—term hospitals during the 12-month period from April 1975 through March 1976. Since their interest lay in studying nosocomial infection, the SENIC researchers excluded women having normal deliveries, patients under 18 years of age, and patients in psychiatric, burn, and other special units. Thus, substantive results from the original index are generalizable to adult medical and surgical patients rather than to the entire hospital population; but careful analysis suggests that any bias resulting from the exclusions is small [11]. Specially trained medical chart reviewers abstracted data on a random sample of approximately 500 general medical and surgical patients in each hospital who were admitted during the study period. The reviewers used standard forms, developed by the Centers for Disease Control, to obtain complete demographic and diagnostic information for each patient. A rigorous system of quality control ensured the highest data quality.

The data set used for the original index excludes 182 patients, or about 0.1 percent of the SENIC sample, for whom no primary diagnosis was recorded. For each remaining patient, however, in addition to the primary diagnosis, the database includes the complete set of all recorded secondary diagnoses. The number of secondaries ranges up to

37, with a mean of 2, although half of all patients have none or only one. The secondary diagnoses are in an arbitrary order, with nothing to distinguish their importance, as is the case with most medical records. All diagnoses are coded according to the hospital adaptation of the International Classification of Diseases (ICDA-8).

We also employed two databases from the National Hospital Discharge Survey [12], as discussed further on, in the Validation section.

METHODOLOGY

In this section we present our strategy for developing a case-mix index. While the strategy is implemented for the dependent variable LOS, it can be used in conjunction with other dependent variables, such as hospital charges. The section includes explanation of the following: grouping of patients, transformation, retransformation, regression and goodness-of-fit, the final model, and sensitivity analysis for the transformation. See also Figure 4, at the end of the section.

GROUPING

We believe that in the absence of a costly assessment of severity made at admission, the only information about the severity and complexity of an illness which is both meaningful and readily available comes from the list of diagnoses on the discharge summary. However, the number of possible diagnoses is too large to handle without collapsing them into categories.

We adopted the major diagnostic category (MDC) system [2] developed originally as the first step toward the DRGs, which collapses the ICDA-8 codes into 83 MDCs based on clinical similarities.[2] (An analysis of the categories determined by our index suggests little difference between ICDA-8 and ICD-9-CM codes for our purposes here.)[3] Because of the very small numbers of patients in a few MDCs, however, we combined MDCs further during the development of our case-mix index. We made these combinations on the basis of clinical similarities, subject to the requirement that for primary diagnoses each MDC should have a sample of at least 100 patients, and for secondary diagnoses at least 450. In addition, we added a new category for trauma patients, whose code begins with the letter E. The result was that primary diagnoses are classified into 73 categories, and secondary diagnoses into 42. These categories are defined in Appendix I. The remainder of this article considers diagnoses only within the context of these collapsed MDCs.

TRANSFORMATION

For patients in our data, LOS ranges from 1 to 360 days, with a mean of 9.4 and a standard deviation of 9.2 days. Figure 1, which displays the distribution, shows clearly its large positive skewness. This suggests that the use of (untransformed) LOS as a dependent variable in regression models should be avoided, because statistical tests which require the assumption of normality would be highly suspect.

A second problem in developing our case-mix index was to account for deaths. LOS clearly does not measure the same attribute for the nearly 4 percent of admissions who die in the hospital as for those who are discharged alive. Simply ignoring discharge status would bias the index, in that hospitals where many patients die soon after admission would then have shorter average lengths of stay, and thus would falsely appear to treat patients more quickly, than hospitals where similar patients are kept alive several weeks before they die or are discharged alive after very long stays. Conversely, to discard the deaths would also introduce a bias. If anything, the situation in which many patients die soon after admission suggests either that the hospital has a more severe case mix than a hospital where similar patients are kept alive, or that the hospital is less capable of managing severely ill patients, or some combination of the two. In any case, excluding deaths creates a bias. Thus, it was desirable to retain in our database all the patients who died in hospital, but at the same time to modify their lengths of stay to reflect better the true burden of their illnesses.

We found that we could simultaneously correct for nonnormality and incorporate deaths by transforming the dependent variable, LOS, as follows:

$$\text{TLOS} = \begin{cases} \text{LOS}/(\text{LOS} + 6) & \text{if discharged alive} \\ 1 & \text{if died in hospital} \end{cases}$$

The constant 6 in this formula was determined empirically.[4] Figure 2, which displays the distribution of TLOS, illustrates that this variable is much more nearly normal than LOS. Note that the values of TLOS for live discharges range from .143 (for LOS equal to 1) to .984 (for LOS equal to 360). Thus, setting TLOS = 1 for patients who died in hospital implicitly assigns death a severity greater than that for patients discharged alive after however long the stay. In this way we avoid the biases which deaths would cause if untransformed LOS were the dependent variable. A logarithmic transformation is frequently used by other authors [9], but it does not result in a distribution as closely

Figure 1: Frequency Distribution of Length of Hospitalization

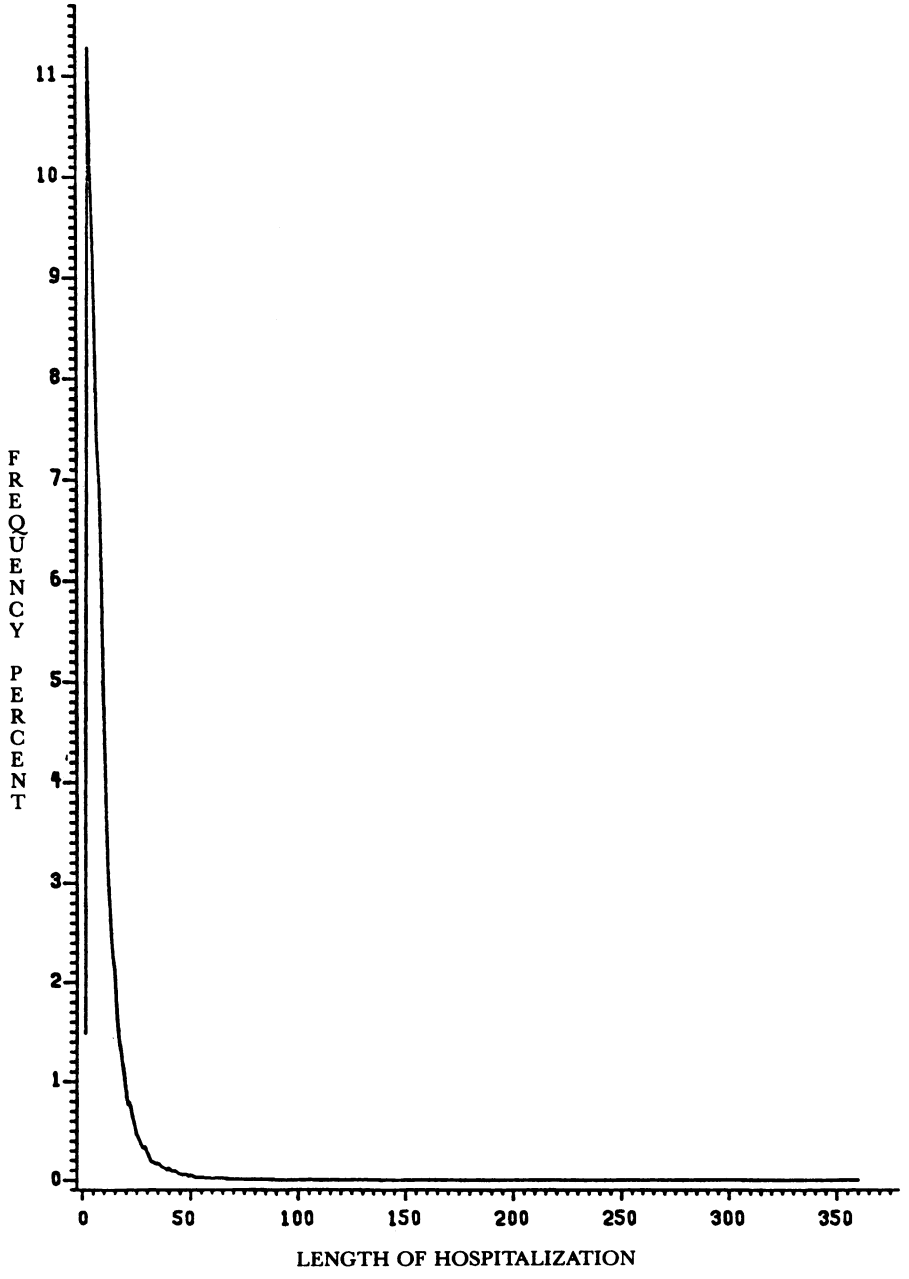
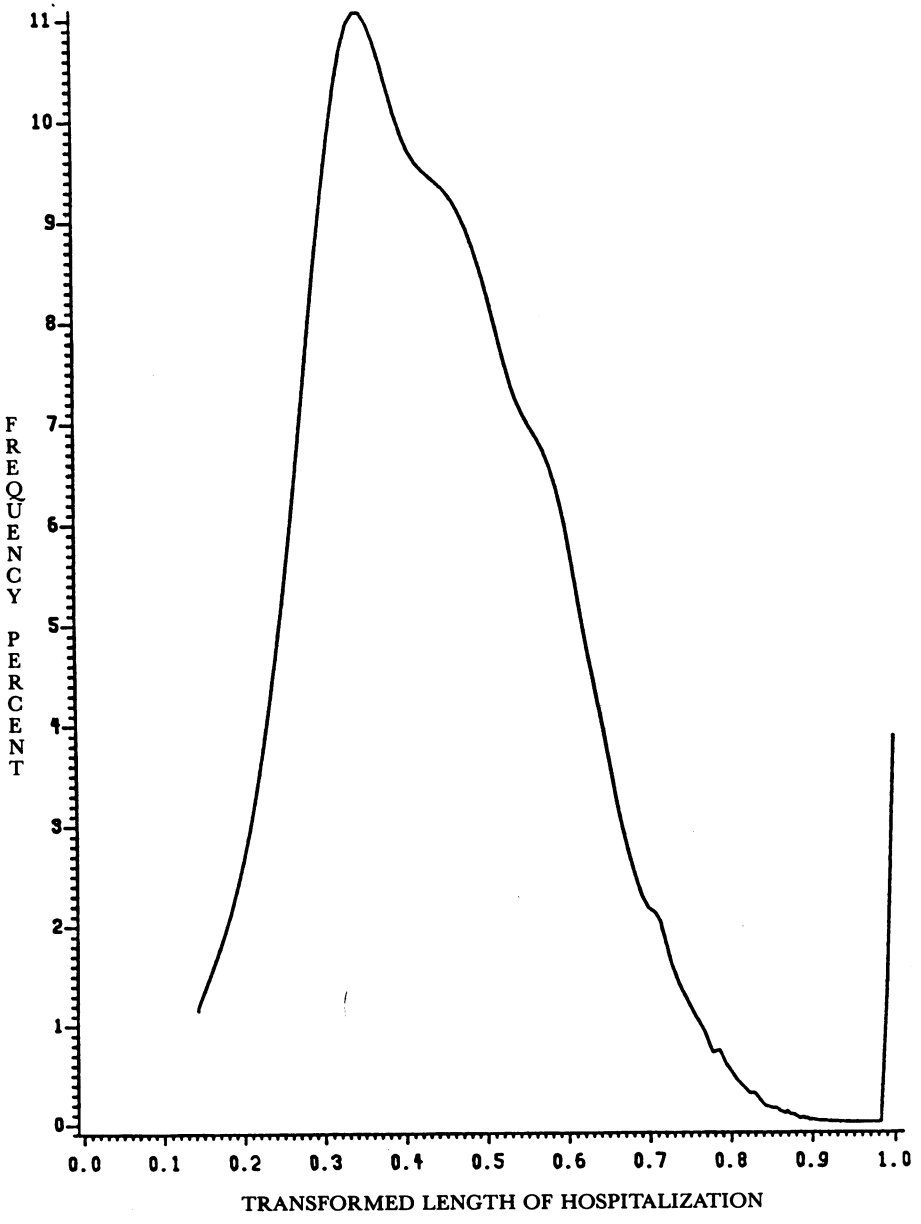


Figure 2: Frequency Distribution of Transformed Length of Hospitalization



normal as ours, and it does not allow deaths to be included as easily as ours.

RETRANSFORMATION

The results of analyses may become harder to interpret if one uses transformed LOS, since the units of TLOS measurement are not actual days of stay. Although we can easily perform statistical tests on the transformed scale, we must transform TLOS back into ordinary patient-days in order to estimate the magnitudes of differences between hospitals. To achieve this objective, we have developed a simple retransformation, denoted D_k, which, when based on hospital A, can be interpreted as the additional stay that would have been observed if a patient who stayed k days in hospital A had gone to hospital B instead. The details of this reverse transformation appear in Appendix II.

REGRESSION AND GOODNESS-OF-FIT

Having adopted the approximately normally distributed dependent variable TLOS, we can construct the case-mix index itself. The degree to which case mix explains variation in LOS is measured by the R^2 value, or the proportion of the total variance of TLOS explained by the case-mix index. While the model with the largest R^2 may be regarded as having the closest fit, it is not necessarily the best. One must balance the virtues of a higher R^2 against those of a parsimonious model (having as few independent variables as possible). This is because policymakers who may wish to make decisions regarding LOS—or regarding other policy-relevant hospital outputs such as unnecessary surgery, reimbursement, use of antibiotics, and the like—must have a case-mix adjustment available to them which is easy to implement and interpret, and therefore is reliant on as few variables as possible.

Our initial attempt at creating a case-mix index used a very simple model: we predicted each patient's TLOS as the mean TLOS for all patients in the same MDC. Stated in technical terms, we regressed TLOS on dummy variables representing the primary MDCs as the only independent variables. The results of this model were not very good ($R^2 = .185$). Next we added a dummy variable indicating whether or not the patient had any surgical procedure. The results of this model were better, but still not the best achievable ($R^2 = .228$). We continued by adding to our regression model other independent factors such as age or secondary diagnoses, and modifications or interactions of factors already in the model. We evaluated their statistical significance by using standard F-tests, but this was not the sole criterion for

deciding whether or not a new factor should be retained; the two goals—a parsimonious model and a large R^2—were considered directly. Examples of the regressions we used showing the variables we tested, and the corresponding degrees of freedom, are in Appendix III.

THE FINAL MODEL

Figure 3 presents the final model used for our case-mix index. In summary, it includes the following factors:

Primary Diagnosis. This factor comprises 73 categories as previously defined, distinguished by dummy variables.

Secondary Diagnoses. Dummy variables represent the 42 categories of secondary diagnoses previously defined. For each patient, the effect of each secondary diagnosis is summed. We found, however, that when a patient has several secondary diagnoses, the importance of each of them is less than when there are only one or two; each additional secondary diagnosis contributes marginally less to LOS than the one before. To correct for the diminishing effects of multiple secondary diagnoses, we included the square of the patient's total number of secondary diagnoses as an additional variable (with a negative coefficient).

Age. We found that the best way to account for age in the model was to use a piecewise-linear function. This was accomplished by including, in addition to age per se, a variable equal to (age - 60) for patients aged over 60 years and equal to 0 for younger patients.

Sex. This, as measured by a dummy variable, emerged as a relatively important factor for some primary diagnoses.

Surgery. This dummy variable indicates whether or not the patient had at least one surgical procedure.

Time Under Anesthesia. We found this to be extremely useful in predicting the stays of surgical patients. We included both linear and quadratic terms in the model.

Interactions. By comparing F-statistics from different regressions, we found the following to be important: the three two-way interactions of primary diagnosis by surgery, primary diagnosis by time under anesthesia, and secondary diagnosis by surgery; and the two three-way interactions of primary diagnosis by surgery by age and primary diagnosis by surgery by sex. The final model does not include any other interactions.

Figure 3: The Case-Mix Index Model

E (TLOS) = a_{ij} + b_{ij} (sex) + SUM(c_k d_{jk}) - e_j (no. of secondaries)^2
+ f_{1ij} (age) + f_{2ij} (max [0, age - 60]) + g_{1i} (time under anesthesia) + g_{2i} (time under anesthesia)^2

where:

a_{ij} = effects of primary diagnosis category i (i = 1, . . . , 73) and surgery category j (j = 1 if yes and 2 if no).

b_{ij} = regression coefficient for sex.

c_k = number of secondaries in category k.

d_{jk} = effect of any secondary in category k.

e_j = regression coefficient for (no. of secondaries)^2.

f_{1ij} f_{2ij} = regression coefficients for age.

g_{1i} g_{2i} = regression coefficients for time under anesthesia.

	DF	SS	MS	F	R^2
Model *	798	1152.2	1.444	72.0	.406
Error	83885	1682.5	0.020		
Total	84683	2834.7			

* Since the model has been presented in the compact "cell means" form, intercept and interactions are automatically included. For example, the terms a_{ij}, 146 in number, include what might more conventionally be shown as an intercept; 72 main effects for primary diagnosis; a main effect for surgery; and 72 primary-surgery interaction effects. The other groups of terms can similarly be seen to provide the remaining main and interaction effects as listed in the text. The actual degrees of freedom depend on the data set: for example, if a particular primary-surgery combination does not occur in the data, one degree of freedom is lost; see Appendix III.

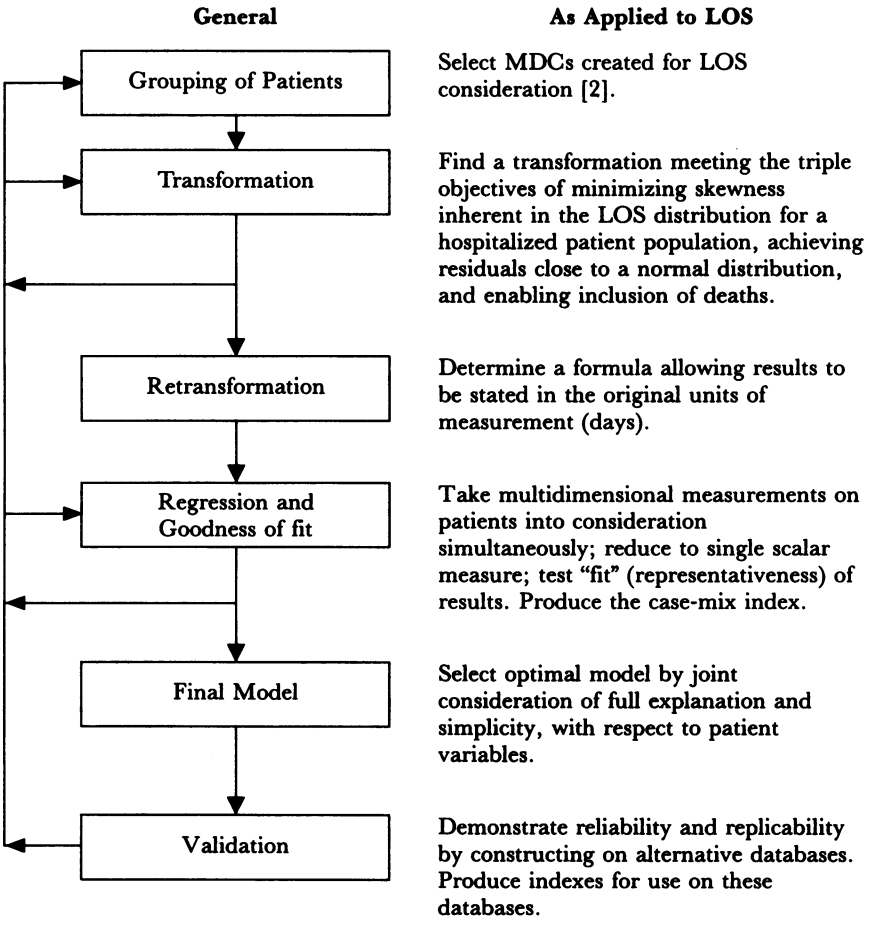
The regression of TLOS on the factors of our final model (Figure 3) yielded an R^2 of 0.406 for the case-mix index. The only way we found to produce a statistically significant increase in this value was to add the interaction between primary and secondary diagnoses. However, the large number of primary-secondary combinations (73 × 42 = 3,066) increases the number of model degrees of freedom to nearly 4,000 while increasing the R^2 only slightly. Thus, we conclude that the index we have chosen, which omits the primary by secondary interaction but uses only 798 model degrees of freedom, is more reasonable.

The network in Figure 4 summarizes the general methodology, along with the application to LOS.

VALIDATION OF THE CASE-MIX INDEX

The value R^2 = 0.406 obtained for the case-mix index is an indication of the quality of the model, but it is also important to determine how

Figure 4: Strategy for Constructing the Case-Mix Index



much of the variance of TLOS the model explains on a separate data set—one not used for development of the index. Anticipating this need, we developed our original index using only a random 50 percent subsample of our data (the base sample, containing 84,684 patients), reserving the other 50 percent (the holdout sample) as an independent data set to use for validation.

For our initial validation procedure we used the case-mix index (i.e., transformation, variables, and regression procedure) to predict

Table 1: Cross-Validation Results—Values of R^2

	Applied to Original Sample Data	*Applied to Holdout Sample Data*
Case-mix index model developed from original sample data	.406	.391
Case-mix index model developed from holdout sample data	.389	.407

Let T_{ik} be the value of TLOS, and X_{ik} the vector of characteristics (diagnoses, age, etc.), for the kth patient in the ith sample ($i = 1$ for the original sample, 2 for holdout). Let M_i be the mean in the ith sample, and let $t_i(x)$ be the predicted TLOS for a patient with characteristic vector X, based on regression analysis using the data of the ith sample. Then the entry in row i and column j of the above table is:

$$R^2_{ij} = 1 - \frac{SUM_k\,[T_{jk} - t_i(X_{jk})]^2}{SUM_k\,[T_{jk} - M_j]^2}$$

the transformed lengths of stay for the patients in the holdout sample, and calculated the amount of (holdout sample) variance explained by our original model. We then computed a second case-mix index (the holdout index) using the case-mix methodology previously discussed but using the data from the holdout sample. We used the resulting new holdout index to predict transformed lengths of stay in the original sample and computed an R^2 value. The results of this cross-validation-like procedure are shown in Table 1. It indicates that while the value of R^2 for the original index is almost 41 percent, we might expect a reduction in variance (R^2) of about 39 percent when applying the coefficients to separate data. The ratios of variance explained in the alternative 50 percent sample to that in the sample utilized for construction are both about 96 percent. We consider this to be an excellent validation of the method.

In order to examine the proportion of variance in LOS that could be explained using our method on a completely external data set, we acquired the 1976 National Hospital Discharge Survey (HDS) [12]. Differences between the HDS and SENIC databases presented certain problems. Since the HDS did not exclude the patients that SENIC did, those exclusions had to be made in the HDS sample to render it compa-

rable. Also, a key variable, time under anesthesia, is not included in the HDS. Therefore, this validation study was restricted to nonsurgical patients. Using nonsurgical patients from the 50 percent SENIC subsample, the R^2 value for the case-mix index is 0.316. The proportion of variance explained by applying the case-mix index to the HDS database in 0.314. We feel that this is an excellent external validation of our case-mix index, especially considering that the care taken in data collection and editing was not as great for the HDS as it was for SENIC.

For a further, final validation of our index, we obtained the 1982 National Hospital Discharge Survey (HDS) [12]. Making the same adjustments as used for the 1976 HDS, and recoding ICD-9-CM codes into ICDA-8 codes, we again calculated the R^2 value for our case-mix index on the 1982 HDS, yielding a value of 0.301. Thus, even on a database taken from a population 6 years later, including all trend changes and a different diagnostic coding scheme, we achieved almost the same proportion of variance explained as in the original database—a ratio of 95 percent. This result strongly suggests that our case-mix adjustment strategy, and the LOS case-mix index in particular, is robust with respect to database employed, categories for patients, and coding scheme utilized.

AN APPLICATION

To give the reader a better appreciation of how the case-mix index can be used to address a policy question, we provide an example. We observed that patients in the SENIC database who were admitted to voluntary hospitals stayed 1.3 days longer on the average (ALOS) than patients in investor-owned (I-O) hospitals (9.6 days versus 8.3). We wanted to determine if this difference in ALOS, which is statistically significant ($p < .01$), could be explained by differences in case mix between voluntary and I-O hospitals. Using our case-mix index, we replaced ALOS by the average residual TLOS (AR)—defined as the average of (actual TLOS - predicted TLOS) over all patients in the hospital—and then regressed it on an indicator variable for I-O hospitals (= 1 for I-O, = 0 for voluntary). (The residual TLOS is the index value used for a patient and AR, the average residual TLOS, is the index value used for a hospital.) As shown in column 1 of Table 2, the difference of -0.0125 in the transformed scale is still statistically significant ($p = .016$); and since the sign is negative, we conclude that the

Table 2: Regressions Illustrating Differences in ALOS Between Voluntary and Investor-Owned (I-O) Hospitals

Independent Variable and Definitions	*Mean*	*Regression Coefficients for AR* (1)	(2)
(Intercept)		X*	X
Indicator of I-O hospital	.13	-1.25	X
LOGBEDS (logarithm of total bed count of hospital)	5.29		.72
Occupancy (percent)	72.42		X
Proportion of patients covered by†			
Private insurance	.52		X
Medicaid	.06		8.61
Medicare	.34		X
Indicator variables for regions†			
New England or E.S. Central	.16		-1.97
S. Atlantic or E.N. Central	.30		-1.30
W.N. Central, W.S. Central, or Mountain	.22		-3.07
Pacific	.13		-7.80
Indicator variables for SMSA size†			
50,000-249,999	.11		1.60
250,000-1,000,000	.20		1.08
1,000,000 or more	.35		1.54
R_2		.016	.572

* X indicates a variable in the model but not statistically significant at the .05 level. Only those coefficients (times 100) significant at $\alpha = .05$ are shown.
†Since the indicator variables for any set of mutually exclusive and exhaustive categories sum to unity, one variable in each set must be omitted in order to prevent linear dependence in the data matrix. Accordingly, we have omitted the indicator variables for "other" method of payment, for the Middle Atlantic region, and for "non-SMSA" (under 50,000).

voluntary hospitals still have the longer average LOS, even after adjustment for case mix. Using the D_{10} retransformation, however, we find that the difference in the original scale has been reduced by more than half, to .5 rather than 1.3 days.

Next, column 2 of Table 2 shows that when we added hospital-level variables such as region and insurance to the regression, the difference between voluntary and I-O hospitals is no longer statistically significant ($p = .57$). Thus, the difference in average length of stay between voluntary and I-O hospitals is completely explained by hospital characteristics and case mix. Fully three quarters of this difference

is explained by case mix alone. A more complete set of regressions, testing for the significance of other explanatory variables and interactions, shows this result to be robust [13].

DISCUSSION

In this article we have described the strategy for constructing an index which incorporates an adjustment for case mix; our particular application produced an index for comparisons of average lengths of stay between individual hospitals or groups of hospitals. We have implemented the construction on three distinct databases. We have examined the quality of the implement indexes with such criteria as R^2, and have performed both internal and external validation procedures. Our indexing approach uses only standard, readily available, archival patient-level variables, with the exception of "time under anesthesia." This variable is highly correlated with other measures of the complexity of surgical procedures, such as operation time and hernia equivalents [14], and is valuable in explaining the variance in the stays of surgical patients. We again point out that the strategy for developing the index might be used to construct particular indexes employable by government or other third-party payers who wish to compare hospital performance in managing medical problems while controlling for case mix or other variables thought to influence LOS, reimbursement, or other hospital output measurements.

This case-mix index has certain advantages over other indexes or classification methods, including broad applicability, independence of database and coding scheme, and the statistical and interpretation features detailed above. While the original index does not contain a direct and comprehensive measure of severity, this is due to the lack of an objective measure in the databases used, an omission which is by no means unusual. Most national or regional databases on which other indexes or patient classification schemes have been developed also lack such a measure. Whenever such information is available, however, it can easily be incorporated into the case-mix index by adding the (combined) severity score as an independent variable.

We demonstrated an important application of the index by noting that when we adjust for case mix and certain hospital variables, the difference in ALOS between voluntary and investor-owned hospitals is no longer statistically significant. This result should be compared to crude (unadjusted) results indicating a large impact due to ownership. We are now using the case-mix index to answer research questions

involving the impact on ALOS of regional differences, medical school affiliation, and other policy-relevant factors.

Our method can also be used in conjunction with DRGs. While we do not attempt to discuss the issue of prospective reimbursement in detail, it is clear that in considering this problem one must adjust carefully for the characteristics of patients in the hospital. Suppose we had total patient charges available for use as our dependent variable. We could apply our index construction strategy and then compare average charges between individual hospitals or groups of hospitals after adjustment for case mix. The potential for improvement of the recently legislated prospective reimbursement plan is illustrated by the following preliminary results. We classified some patients in our database into DRGs (ICDA-8 version), using rules inherent in the AUTOGRP program [15]. We then measured the proportion of variance of TLOS explained by our case-mix index within each DRG. The assumption behind the use of DRGs for reimbursement is that the patients within any DRG are homogeneous; therefore, additional variance explained by our index is evidence of lack of homogeneity of the DRGs. Although we reduced variation very little in some DRGs, we obtained large reductions—of up to 28 percent—in several DRGs. The legislative, economic, and political consequences of such differences are meaningful and important. The reader should keep in mind, however, that these results are preliminary, and do not use (average) costs as a dependent variable.

Another area of potential application is drug usage. Many researchers are concerned with the use of antibiotics and other medications in hospitals [16]. Suppose we defined a dependent variable reflecting the frequency of total drug administrations, or administrations of a specific class of medications such as antibiotics. We could transform this dependent variable appropriately and obtain a model similar to that outlined above for the case-mix index. The result would be a drug or medication index which, again, would facilitate comparisons of hospitals.

These are just two examples of the possible use of the methodologic approach of our case-mix index in other situations. The results we obtained with this index in studying hospital variations in length of stay suggest that such an approach can be valuable in several other areas.

APPENDIX I

FINAL LISTS OF COLLAPSED MDCS FOR PRIMARY AND SECONDARY DIAGNOSES

Original MDCs	*Name*	*Collapsed MDCs for Primary Diagnosis*	*Collapsed MDCs for Secondary Diagnosis*
1	Infectious diseases	1	1
2	Malignant neoplasms of digestive system	2	2
3	Malignant neoplasm of respiratory system	3	2
4	Malignant neoplasm of skin	4	2
5	Malignant neoplasm of breast	5	2
6	Malignant neoplasm of female genital organs	6	2
7	Malignant neoplasm of male genital organs	7	2
8	Malignant neoplasm of urinary system	8	2
9	Malignant neoplasm of other and unspecified sites	4	2
10	Neoplasm of lymphatic and hemopoietic tissue	9	2
11	Benign neoplasm of female genital organs	10	3
12	Benign neoplasm of other sites	11	3
13	Diseases of thyroid and other endocrine glands	12	4
14	Diabetes	13	5
15	Nutritional and other metabolic diseases	14	6
16	Diseases of blood and blood-forming organs	15	7
17	Psychoses not attributed to physical conditions	16	8
18	Neuroses	17	8
19	Alcoholic mental disorder and addiction	18	8
20	Other mental disorders	16	8
21	Diseases of central nervous system	19	9
22	Diseases of peripheral nervous system	20	10
23	Diseases of eye	21	10
24	Diseases of ear and mastoid process	21	11
25	Hypertensive heart diseases	22	12
26	Acute myocardial infarction	23	13

Continued

Appendix I: Continued

Original MDCs	*Name*	*Collapsed MDCs for Primary Diagnosis*	*Collapsed MDCs for Secondary Diagnosis*
27	Ischemic heart diseases except AMI	24	13
28	Arrhythmia and slowed conduction	25	14
29	Heart failure	26	15
30	Carditis, valvular, and other diseases	27	16
31	Cerebrovascular diseases	28	9
32	Diseases of vascular system	29	17
33	Pulmonary embolism	30	18
34	Phlebitis and thrombophlebitis	31	18
35	Hemorrhoids	32	19
36	Hypertrophy of tonsil and adenoid	33	11
37	Acute URI and influenza	33	11
38	Other diseases of upper respiratory tract	21	11
39	Pneumonia	34	20
40	Bronchitis	35	21
41	Asthma	36	21
42	Other lung and pleural diseases	37	21
43	Diseases of oral cavity, salivary glands	21	11
44	Gastric and peptic ulcer	38	22
45	Upper GI diseases except gastric and peptic ulcer	39	22
46	Appendicitis	40	23
47	Hernia of abdominal cavity	41	24
48	Enteritis, diverticula, and functional disorder of intestine	42	25
49	Diseases of anus	43	19
50	Miscellaneous diseases of intestine and peritoneum	44	25
51	Diseases of liver	45	26
52	Diseases of gall bladder and bile duct	46	23
53	Diseases of pancreas	47	23
54	Diseases of kidney and ureter	48	27
55	Urinary calculus	49	27
56	Cystitis and other urinary diseases	50	28
57	Diseases of prostate	51	19
58	Diseases of male genital organs	52	19
59	Diseases of female genital organs	53	29
60	Diseases of breast	54	30
61	Abortion	55	29
62	Obstetrical diseases of antepartum and puerperium	56	29

Continued

Appendix I: Continued

Original MDCs	*Name*	*Collapsed MDCs for Primary Diagnosis*	*Collapsed MDCs for Secondary Diagnosis*
63	Normal delivery	57	29
64	Delivery with complication	57	29
65	Diseases of skin and subcutaneous tissue	58	30
66	Arthritis	59	31
67	Derangement and displacement of intervertebral disc	60	32
68	Diseases of bone and cartilage	61	32
69	Other diseases of musculoskeletal systems	62	32
70	Congenital anomalies	63	33
71*	Normal mature births	—	—
72*	Certain diseases and conditions peculiar to newborn infants	—	—
73	Symptoms and signs referable to nervous, respiratory, and circulatory systems	64	34
74	Symptoms and signs referable to GI and urinary system	65	35
75	Miscellaneous signs, symptoms, and ill-defined conditions	66	36
76	Fractures	67	37
77	Dislocation and other musculoskeletal injury	68	37
78	Internal injury of cranium, chest and other organs	69	37
79	Open wound and superficial injury (H-ICDA-2 Codes 910-939.1)	69	38
	(H-ICDA-2 Codes 870-897.9)	70	38
80	Burn	69	38
81	Complication of surgical and medical care	71	39
82	Adverse effects of a certain substance	72	40
E-codes	Supplementary classifications of external cause of injury	73	41
Y-codes	Supplementary classifications (not elsewhere classified)	73	42

*For neonates only; thus, these patients are excluded.

APPENDIX II

THE REVERSE TRANSFORMATION

Suppose we wish to compare two hospitals (or groups of hospitals), say A and B. We apply the case-mix index to all patients in both hospitals, and obtain a residual TLOS for each patient. The next step requires averaging these residuals to obtain the mean residual transformed LOS for each hospital, $RTLOS_A$ and $RTLOS_B$. An estimate of the difference in LOS between the two hospitals on the transformed scale, controlling for case mix, is $d = RTLOS_B - RTLOS_A$. The problem is to derive a similar estimate, say D, on the LOS scale.

If a patient in hospital A has a transformed stay equal to T, we estimate that the same patient would have had a transformed stay equal to $(T + d)$ in hospital B. On the LOS scale, the patient in hospital A actually stayed $L = 6T/(1 - T)$ days (if he/she had not died). A reasonable estimate of the same patient's stay in hospital B is $6(T + d)/(1 - T - d)$ days. Some algebra then reveals that the estimated difference is $6\,d/[(1 - T)(1 - T - d)]$ days. The required value of D is the average of the estimated differences over all patients in hospital A. We interpret D as the additional stay which would be observed if an "average" patient in hospital A had gone instead to hospital B.

The method just given requires a separate computation for each patient, if one wants a comparison averaged over all patients. A quick and simple alternative comparison involves standardizing to the "typical" patient who stays, say, 10 days. A patient in hospital A who stayed 10 days would have $TLOS = 5/8$. If this patient had gone to hospital B, we would estimate an additional stay of $D_{10} = 6d/[(1 - 5/8)(1 - 5/8 - d)] = 128d/(3 - 8d)$ days.

APPENDIX III

DEGREES OF FREEDOM IN THE MODELS CONSIDERED

*Term**	*Model 1*	*Model 2*	*Model 3*	*Model 4*
Intercept	1	1	1	1
Primary diagnosis	72	72	—	72
a_{ij}		73	145	—
f_{1ij} * primary			—	73
f_{1ij} * a_{ij}			146	—
f_{2ij} * primary			—	70
f_{2ij} * a_{ij}			140	—
b_{ij} * primary			—	69
b_{ij} * a_{ij}			135	—
g_{1i} * primary			—	73
g_{1i} * a_{ij}			73	—
g_{2i} * primary			—	73
g_{2i} * a_{ij}			73	—
Secondary * Primary			—	2,986
d_{jk}			84	—
e_j			2	—
SUM Secondary * Primary			—	73
Reduced model *DF*	72	145	798	3,489
Model degrees of freedom unused due to combinations of variables not present in the data	0	0	17	87
R^2	.1847	.2283	.4065	.4330

*See Figure 3 for definition of terms. Users may request a printout with coefficients for Model 3 from the authors.

ACKNOWLEDGMENTS

We wish to acknowledge the help and support of researchers at the Centers for Disease Control who were responsible for the original SENIC project, and who have always responded to requests for help with the current LOH project. These include, but are not limited to, the SENIC project chief, Dr. Robert W. Haley, now at the University of Texas (Dallas), and its chief statistician, Dr. David H. Culver. We would also like to thank two anonymous referees and the editor for their helpful comments. Finally, almost all work on the LOH project is the result of a team effort; thus, we wish also to acknowledge our fellow researchers, Dr. Lisa Kammerman, Mr. Steve Wilcox, and Dr. Adela Sanz.

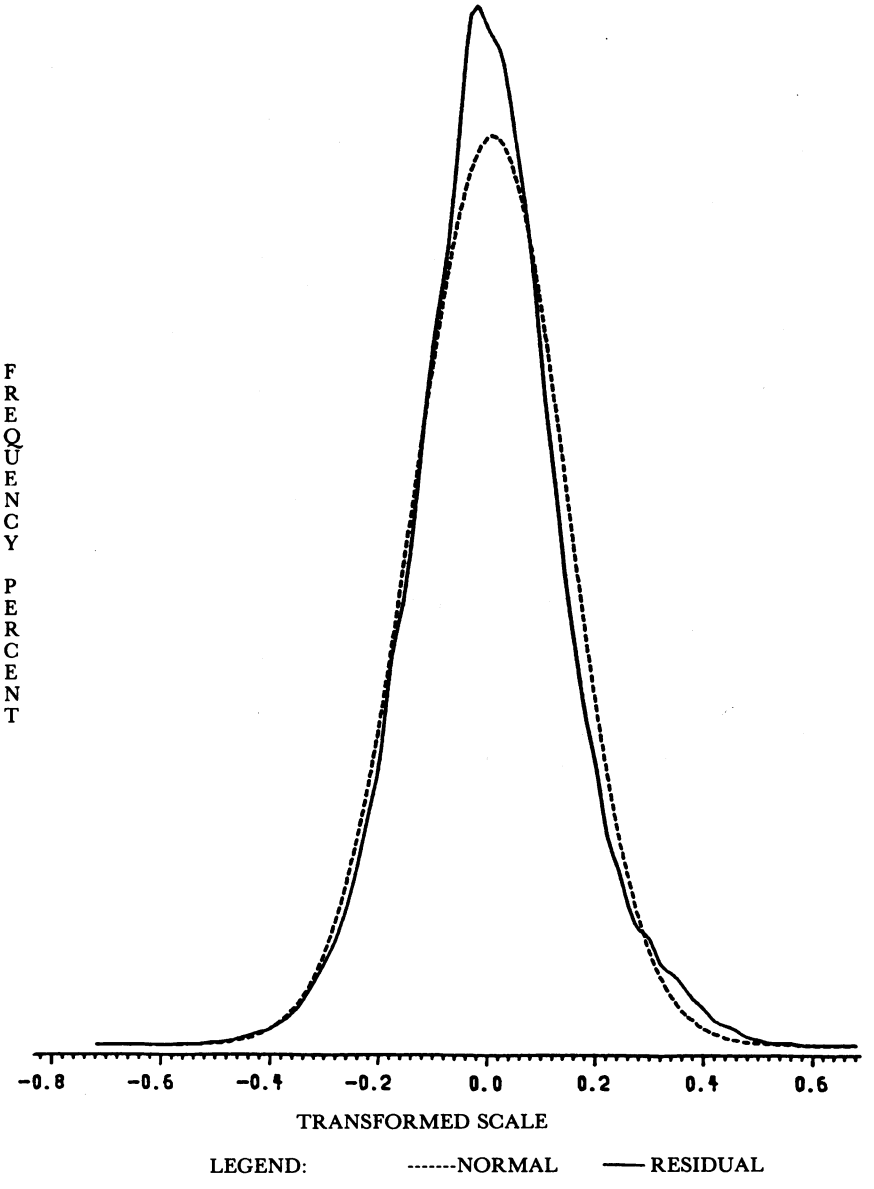
NOTES

1. The "ecological fallacy" (or "aggregation bias") occurs when relationships found at one level of analysis (e.g., hospitals) are assumed to hold at another level (e.g., patients)—for example, if longer average stays for hospitals with high proportions of Medicare patients were thought to imply automatically longer average stays for Medicare patients.
2. We chose to work with the MDCs since they are (1) generated based on clinical considerations [2], (2) well defined [2], (3) readily available [2, 12], and (4) already the basis for a well-known classification system [2]. In fact, any system of patient categories, whether based on etiology, organ system, or other criteria for clinical homogeneity, would be a reasonable candidate for the basic building blocks for the index. The key research issue in selecting a grouping is: "Do the categories in the grouping allow for a standardization of patients which is sufficiently detailed to adjust for case mix when comparing hospitals, yet not too complex to use in calculations?"
3. The reasoning of Note 2 also applies to code differences. The system of codes used, whether ICDA-8, ICD-9-CM, or another, is less important than whether they provide sufficient detail to classify patients into the categories for grouping. In the version derived here, this is even less of a problem, since the codes are used solely to classify patients into MDCs.
4. To measure the sensitivity of the distribution to the choice of the constant 6, we used a random subsample of 8,500 patients for a series of regression analyses, identical except that in each analysis the dependent variable was defined with a different constant. The analysis utilized three different criteria for measuring normality of the residuals: the Kolmogorov-Smirnov statistic, skewness, and excess of kurtosis. For each criterion, a value close to zero is indicative of a normal distribution. According to these criteria, the constant 5 would be optimal for normalizing the residuals. To maximize R^2, however, we found that a constant between 6 and 8 would be optimal. The value 6 was then adopted as a compromise. This seems satisfactory since the residuals from the resulting case-mix index were still very nearly normal (Figure 5). To satisfy the assumptions of the regression model which are to be used in exploring the impact of hospital variables on LOS after adjusting for case mix, it is necessary for the residuals to be normally distributed. In addition, even in validation runs on alternative databases (under "Validation of the Case-Mix Index"), results were not sensitive to small changes in the constant selected, once adjusted for the overall ALOS in the database utilized.

REFERENCES

1. Hornbrook, M. D. Hospital case mix: Its definition, measurement and use, Part II. Review of alternative measures. *Medical Care Review* 39(2):73, 1982.
2. Fetter, R. B., et al. Case mix definition by diagnosis-related groups. *Medical Care* 18(2):Supplement, 1980.
3. Fetter, R. B., et al. The New ICD-9-CM Diagnosis Related Groups

Figure 5: Frequency Distribution of Case-Mix Index Residuals Compared with a Normal Distribution



Classification Scheme, Final Report. Health Systems Management Group, School of Organization and Management, Yale University, New Haven, 1982.
4. Commission on Professional and Hospital Activities. *Hospital Adaption of ICDA (H-ICDA), Second Edition. Ann Arbor, MI: CPHA, 1973.*
5. *International Classification of Diseases—Clinical Modification*, Second Edition, Vol. 3. U. S. Department of Health and Human Services, Public Health Service, Health Care Financing Administration, Washington, DC: Government Printing Office, 1980.
6. Horn, S. D., P. D. Sharkey, and D. A. Bertran. Measuring severity of illness: Homogeneous case mix groups. *Medical Care* 21(1):14, 1983.
7. Gonnella, J. S., D. Z. Louis, and J. J. McCord. The staging concept—an approach to the assessment of outcome of ambulatory care. *Medical Care* 14(1):13, 1976.
8. Young, W. W., R. B. Swinkola, and D. M. Zorn. The measurement of hospital case mix. *Medical Care* 20(5):501, 1982.
9. Pettengill, J., and J. Vertrees. Reliability and validity in hospital case mix measurement. *Health Care Financing Review* 4(2):101, 1982.
10. Quade, D., et al. The SENIC sampling process: Design for choosing hospitals and patients and results of sample selection. *American Journal of Epidemiology* 111(5):486, 1980.
11. Shachtman, R. H., et al. An analysis of potential bias due to using the SENIC database for length of hospitalization studies. LOHNOTE 24, Department of Biostatistics, University of North Carolina, 1983.
12. U. S. Department of Health and Human Services. Public use data tape documentation—*National Hospital Discharge Survey 1970–1978, 1981, 1982.* Hyattsville, Maryland, 1981-1984.
13. Freund, D., et al. Analysis of length of stay differences between investor-owned and voluntary hospitals. *Inquiry* 22(1): 33, Spring 1985.
14. Hughes, E. F. X., et al. Surgical work loads in a community practice. *Surgery* 71(3): 315, 1972.
15. Mills, R. E., et al. AUTOGRP: An interactive computer system for the analysis of health care data. *Medical Care* 14 (7):503, 1976.
16. Broadie, D. C., and W. E. Smith. Constructing a conceptual model of drug utilization review. *Hospitals* 50(6):143, 1976.