

Are the diagnosis-related group case weights compressed?

by Kenneth E. Thorpe, Shan Cretin, and
Emmett B. Keeler

One problem noted recently with the diagnosis-related group payment system is that the distribution of Medicare case weights and case-mix indexes are compressed; that is, the payment rates for high-cost procedures are too low and those for low-cost procedures are too high. Despite the attention

compression has received, there are no direct estimates of its magnitude or importance. Presented in this article are an empirical test for compression and a suggestion for a simple correction to decompress the relative prices.

Introduction

The Medicare case payment system pays hospitals a predetermined, fixed price for each of 471 different diagnosis-related group (DRG) categories. Hospital payment rates for each DRG are the product of a standardized cost and relative price, or cost weight, established for the DRG category. The payment rates are meant to approximate the average operating cost per case for each DRG. The structure of the relative price schedule and the payment rates for hospitals are likely influenced by the variables used to standardize costs, the averaging procedure used, the accuracy of hospital coding, and by hospital pricing and cost-accounting practices.

One problem noted recently is that the distribution of the Medicare case weights and the case-mix indexes are compressed (Pettengill and Vertrees, 1982; Lave, 1985a); that is, the relative prices of low-cost DRG's are too high and high-cost DRG's are too low relative to their true costs. Compression has generally been distinguished from coding problems with specific DRG categories that would lead to misestimation of some relative weights. Instead, compression has been seen as a systematic misestimation of weights that depends not on clinical details of a DRG category but on the actual numerical value of the weight for each DRG.

Three factors may account for compression in the DRG prices. First, the cost-estimating algorithm used to establish the relative DRG case weights may contribute to compression. In this estimation procedure, routine and special care per diem costs for each DRG are assumed constant (Lave, 1985b). It is likely, however, that nursing time and, hence, costs per discharge vary across the DRG categories.

Moreover, there is significant variation in costs among patients in intensive care units that is not captured by lengths of stay associated with special care (Wagner and Draper, 1984). Seriously ill patients would incur greater resource use than those who are

simply under observation. Thus, it is possible that the special care costs assigned to DRG's whose patients generally receive active treatment will be low relative to the weight assigned to those DRG's whose patients are often monitored. Failure to account for these differences would lead to compression in the DRG relative prices.

Second, in the past, hospitals have frequently cross-subsidized some ancillary services (Harris, 1979; Office of Technology Assessment, 1983). In particular, reported hospital costs for sophisticated diagnostic procedures may understate actual costs, and those reported for less technical ancillary services may be too high. Such accounting practices might contribute further to compression in the DRG case weights.

Finally, the initial data set used to create the DRG case weights contains many coding and classification errors (Institute of Medicine, 1977). Initial coding errors would bias the estimated case weights toward the average case weight, thus compressing further the structure of the DRG case-weight prices. The important potential role assumed by classification errors in compression was simulated by Pettengill and Vertrees. Compared with existing coding practices, 30 percent additional error would reduce the standard deviation of the case weights from .078 to .068. (Pettengill and Vertrees, 1982; Lave, 1985a).

Existing evidence of case-weight compression

Despite the attention the compression issue has received, there have not been any direct tests of its magnitude, although two attempts to indirectly estimate the extent of compression in the DRG case weights have recently been completed (Lave, 1985a; Cotterill, Bobula, and Connerton, 1986). Lave compared the estimated means and standard errors of the existing Medicare DRG case weights with those calculated using data from two waived States, Maryland and New Jersey. Lave suggests that the distribution of the case weights and case-mix indexes in these two States may be less compressed because of more accurate data coding in these States and because of existing regulatory systems that discourage widespread cross-subsidization across ancillary departments. The results indicated that there was more variation in the distribution of the case weights

Support for this article was provided by the Health Care Financing Administration under the terms of Cooperative Agreement Number 99-C-98489/9-02, as part of the RAND/University of California, Los Angeles Center for Health Care Financing Policy Research.

Reprint requests: Kenneth E. Thorpe, Department of Health Policy and Management, Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115.

in the waived States compared with the Medicare DRG case weights. Lave's results support the belief that the DRG relative prices are compressed.

Cotterill, Bobula, and Connerton (1986) compared the distribution of the original Medicare DRG case weights based on costs with DRG weights calculated using hospital charges. They concluded that the resulting charge-based case weights and case-mix indexes are less compressed when using charge data. They attribute this result to current hospital pricing practices, where the ratio of total charges to total costs tends to be relatively larger for high-weight DRG's and lower for low-weight DRG's. Although use of charge data to calculate the relative DRG weights decompresses the distribution of the case-mix indexes, it is not clear that they more accurately reflect the true distribution of case weights.

Importance of compression

Compression in the Medicare case-mix indexes would bias payment levels to hospitals. Under the current payment formula, a compressed case-mix index would result in hospitals with relatively high case-mix indexes being underpaid and those with lower case-mix indexes being overpaid. Compression could also influence the profitability of individual DRG categories because low-cost cases will, *ceteris paribus*, be relatively profitable, and higher cost cases may be less profitable. Determining whether or not the case-mix index is compressed, then, has important public policy implications. If, indeed, there is compression, then estimating the degree of compression and proposing an appropriate adjustment to the Medicare payment system has a high priority.

The current Medicare payment formula includes factors beyond the case-mix index in determining hospital payment. For example, the indirect teaching adjustment increases the payment to teaching hospitals, the urban-rural adjustment results in higher payments to urban hospitals, and the recently adopted disproportionate-share adjustment generates higher payments to hospitals serving the poor. The question of compression in the case-mix index must be addressed, therefore, in the context of the other payment adjustments in the Medicare prospective payment system. Simply decompressing the Medicare case-mix index would clearly redistribute Medicare payments from low- to high-complexity institutions; it is not so clear that such a redistribution would result in net Medicare payments that better reflect hospital costs. Moreover, it is not clear that a better reflection of variations in hospital costs is desirable. The desirability of such redistribution hinges on the factors responsible for any unmeasured variation in cost per discharge.

Empirical estimation: Case-weight compression

The goals of this article are twofold: to estimate empirically the degree of compression in the charge-

based DRG case weights and to suggest a simple method to decompress the case weights. Once the case weights are adjusted, we shall then compare the performance of compressed and decompressed DRG weights in explaining variation in hospital cost per case.

Our general approach to assessing the extent of compression in the case-mix index is modeled on that of Pettengill and Vertrees (1982). We use hospital-level data to estimate an average cost function using the case-mix index as one of the independent variables. Because of its construction, the case-mix index should be proportional to the estimated average operating costs per discharge. If the index is compressed, however, the estimated coefficient in the cost function would not be proportional. Under the assumption of compression in the case weights, the estimated coefficient on the case-mix index would be more than proportionally related to costs. Moreover, depending upon the correlations among all the exogenous variables in the equation, this bias could spread to other regressors in the system, potentially affecting their estimated relationships with average operating costs. Thus, any bias is problematic because parameter estimates from hospital-cost regressions were used to develop the indirect teaching and disproportionate-share adjustment. We will explore this problem by using several different specifications of the cost function in our analyses.

In the next section, we present in detail the methods used in this study, including a model of compression, specifications of hospital cost functions, a description of the data set, and the statistical approaches employed. Also presented in this article are the results of our analyses and a discussion of these results and their implications for Medicare reimbursement policy.

Methods

Model of case-weight compression

The first problem in estimating case weight compression is to define precisely what is meant by compression. We have already noted that compression in the case weights would result in systematically underestimating the weights of high-weighted DRG's and overestimating the weights of low-weighted DRG's. However, the exact functional form of this systematic misestimation is not uniquely specified by the general arguments for the existence of compression. Before we can quantify the degree of compression, we have to make assumptions about the form of the compression. For the purposes of this analysis, we will assume that the compression in the case weights is such that the distribution of case weights undergoes a linear transformation.

Our model of compression can be formularized as follows. Let W be the distribution of case weights W_i , normalized so that the mean over hospitals $E(W) = 1$ and the standard deviation is $S(W)$. We hypothesize that there exists a true distribution W' with mean $E(W') = 1$ and standard deviation $S(W') > S(W)$. In

the simplest case, W' can be modeled as a linear transformation of W . If W_i is the case weight for the i th DRG, for all i , then

$$W'_i = k(W_i - 1) + 1 \quad (1)$$

In any linear transformation, k is the ratio of $S(W')/S(W)$. In the special case when $k = 1$, the estimated and actual case weights are equivalent. If the existing DRG case weights are compressed, however, this ratio would exceed 1 (Pettengill and Vertrees, 1982).

What happens at the hospital level to the case-mix index using these two different sets of weights? Let C_j be the case-mix index for hospital j constructed from W , the compressed DRG weights, and C'_j be the case-mix index constructed using the true weights, W' . Given the current construction of the Medicare case-mix values, then

$$C_j = \sum W_i F_{ij} \quad (2)$$

where F_{ij} is the proportion of all discharges at hospital j in DRG i , so that the F_{ij} sum to one at each hospital. Substituting equation (1) into equation (2) and its primed counterpart, we see that equation (1) remains true when C is substituted for W :

$$C'_j = k(C_j - 1) + 1 \quad (3)$$

Given this formulation and after adjusting for other exogenous factors influencing costs, the C'_j should be proportional to the average cost per case at hospital j . If the current case-mix index is used to predict cost per discharge in a statistical cost function, estimation of k (the degree of compression) allows a direct test of the existence and extent of compression in the current DRG case weights.

Our measure of the degree and importance of compression remaining in the latest Medicare relative weights will be derived from the estimated coefficient on the case-mix index in a hospital cost function. The estimated coefficient could be affected in two ways: by compression in the case-mix index and by relevant variables omitted from the equation. To isolate compression in the case-mix index from any omitted variable bias in the model, all relevant explanatory variables should be included in the estimated cost function. These variables and their definitions will be discussed later.

Specification of hospital cost functions

Our proposed test of compression will be influenced by how we choose to estimate costs, our dependent variable, and by which other exogenous factors we include with the case-mix index in our independent variables. Following the approach used by Pettengill and Vertrees (1982), we chose operating cost per discharge as the appropriate measure of hospital cost. We have also used many of the same explanatory variables. We have also added independent variables

that have been suggested subsequently as important factors in determining hospital costs. In actually carrying out our analyses, we will employ models using from 4 to 19 independent variables. Changing the explanatory variables included in the model will allow us to observe the degree to which omitting relevant independent variables contributes to apparent compression.

The independent variables used in estimating our hospital cost functions include the following:

- Hospital DRG case-mix index.
- Medicare wage index.
- Hospital bed size.
- Ratio of interns and residents per bed.
- Proportion of days covered by Medicaid.
- Hospital ownership.
- Census region.
- City size.
- Urban-rural location.

We will discuss each of these variables in turn.

Hospitals serve many different types of patients. This heterogeneity requires some standardization for differences in hospital inpatient case mix (Caves and Christensen, 1980; Cowing and Holtmann, 1983). To standardize for variations in hospital case mix, we shall employ the Medicare case-mix index for fiscal year 1986 calculated by the Health Care Financing Administration (HCFA). This 1986 index differs from the previous case-mix index in that the relative DRG prices (weights) were calculated using hospital charges rather than costs, and the calculation was based on more recent (1984) data. Moreover, the charge weights included capital and direct medical education costs, whereas the earlier, cost-based weights did not. The charge-based weights are, indeed, less compressed than those based on 1981 cost data, having a larger standard deviation and coefficient of variation (Table 1). But are these new charge weights still compressed relative to the true weights? Do these new charge weights provide a more valid measure of hospital case mix?

Table 1

Comparison of cost-based and charge-based case-mix indexes

Case mix based on:	Mean	Standard deviation	Coefficient of variation
Charge data	1.028	0.141	0.137
Cost data	1.047	0.136	0.130

NOTE: These summary statistics reflect unweighted case-mix indexes of those hospitals included in the sample. Cost weights and the resulting case-mix index are based on the 1981 MEDPAR data file.

All of our models will include either a compressed or decompressed measure of case mix based on the charge-based case weights. To test the 1986 weights, we constructed our own (bill) weighted average case-mix index for each hospital in the sample. This index reflects true costs better than the actual DRG weights used for payment in the 1984-85 sample years that were based on simpler cost estimation methods and earlier data.

The measure of hospital wages used in the cost

function will be the wage index used by HCFA in determining DRG payment rates (*Federal Register*, Sept. 3, 1985). All of our models will include the wage index.

Although the cause of the relationship of hospital size to hospital costs is still disputed, there is evidence to suggest that bed size does predict costs (Pettengill and Vertrees, 1982). Appropriate payment mechanisms will, indeed, depend on whether this relationship is the result of inefficiencies, diseconomies of scale, or unmeasured complexity of cases at larger institutions.

Teaching hospitals are thought to generate higher costs per discharge than other hospitals (Pettengill and Vertrees, 1982; Sloan, Feldman, and Steinwald, 1983). These additional indirect costs are commonly believed to be created either through the hospital's teaching mission, through more aggressive treatment patterns, through unmeasured differences in the complexity (severity) of patients treated, or all the above. As a consequence, additional tests and ancillary services per case result. To control for these differences between teaching and nonteaching hospitals, a measure of the scope of teaching responsibility—the ratio of interns and residents per bed—will be employed. All models will include this variable.

The volume of low-income and/or Medicare Part A patients treated by the hospital may also affect operating cost per discharge (Prospective Payment Assessment Commission, 1985). These patients could increase average costs if they are more severely ill than other patients within any given DRG category. Ideally, one would want to include some characterization of each hospital's per capita Medicare income level. Because these data are not available, we shall use the proportion of Medicaid days from each hospital as a proxy.

Hospital ownership is also included as a control variable in the cost function. Although the empirical evidence is mixed, systematic differences in efficiency related to ownership could affect cost per discharge (DeAlissi, 1983; Becker and Sloan, 1985).

Regional dummy variables were also included in the cost function. These variables are entered to capture unmeasured variations in costs—resulting from regional differences in length of stay, admitting patterns, and ancillary use—not explicitly included in the cost functions (Mitchell, 1985).

Finally, hospital location was characterized as being in a rural area, a small, medium, or a large metropolitan statistical area city. The city-size variables were used to reduce within-region variations in costs not easily measured by the other regressors. Small area variation in practice patterns and nonwage price differences within each census region could affect cost per discharge. If so, the city-size dummy variables should reflect these differences. In some of our models, a simple urban-rural split is used.

Data sources

The analysis file contained a full year of Medicare bills for each hospital under prospective payment

system (PPS). Because hospitals entered PPS at different points in time, bills during the first year's experience covered portions of Federal fiscal years 1984 and 1985. The accounting cost per Medicare discharge was estimated by combining information from two sources: the 1984 Medicare cost reports and a 20-percent sample of all bills during each hospital's first year of PPS.

The Medicare cost reports were used to estimate hospital-specific ratios of cost to charges for each department and costs per day for routine accommodations and special care accommodations. In order to adjust for inflation between the time of the Medicare cost report and the date on a patient's bill, we inflated per diem costs using the estimated HCFA hospital market basket inflation rates. Direct medical education and capital expenditures were excluded.

Some hospitals for which there were bills in fiscal year 1984 had incomplete or clearly erroneous data in their 1983 cost report. When possible, missing or erroneous cost report data were replaced with an estimate that was based on the average value for other hospitals with similar characteristics. Any admissions for which any part of the estimated cost had to be based on estimated rather than on actual cost report data were flagged.

Estimated accounting costs for each discharge were derived by summing routine accommodation costs, special care accommodation costs, and total ancillary costs. The accommodation costs for an admission were obtained by multiplying the appropriate per diem by the length of stay. The total ancillary costs for an admission were obtained by multiplying charges in each department by the hospital's ratio of cost to charges in that department. A detailed description of construction of the accounting cost variable is included in a report by Newhouse, Cretin, and Witsberger (1987).

Finally, a hospital level file was created for the compression analysis. In this file all the Medicare discharges from each hospital under PPS were aggregated. Using the methods described earlier, estimated costs per Medicare discharge, the DRG case-mix index, and the percent of Medicare bills flagged because of missing or bad cost report data were calculated at the hospital level. Other hospital characteristics—such as the number of beds, the area wage index, the ratio of interns and residents per bed, type of control (public, proprietary, nonprofit community), and the census region—were merged from provider level files obtained from HCFA. The proportion of all admissions to the hospital covered by Medicaid and the city-size variables were obtained from the 1984 AHA Hospital Survey. The resulting file contained observations on 5,015 hospitals.

Table 2 contains a summary of the variables included in the analysis for the hospitals included in our analyses.

Statistical model

The primary problem in modeling average costs per discharge is that the distribution of average costs is

Table 2
Variable definitions, means, and standard deviations

Dependent variable	Mean	Standard deviation
Medicare operating cost per discharge (CPD)	2,493.83	1,005.39
Log (CPD)	7.74	0.381
Exogenous variables		
Medicare case mix (DRG)	1.028	0.140
Log (DRG)	0.018	0.135
Medicare proposed wage index (WAGE)	0.961	0.149
Log (WAGE)	-0.05	0.149
Hospital beds (BEDS)	146.91	147.75
Log (BEDS)	4.55	0.954
Ratio of interns, residents per bed (INT)	0.019	0.069
Log (1 + INT)	0.017	0.054
Proportion of admissions covered by Medicaid (CAID)	0.067	0.070
Log (1 + CAID)	0.082	0.060
Hospital located in urban area with population greater than 1,000,000 (LMA)	0.229	0.421
Hospital located in urban area with population 250,000 to 1,000,000 (MMA)	0.162	0.368
Hospital located in urban area with population less than 250,000 (SMA)	0.097	0.295
Hospital located in rural area (NONMET)	0.512	0.50
Hospital located in New England census division: (NENG)	0.023	0.151
Middle Atlantic (MIDAT)	0.044	0.206
South Atlantic (SOATL)	0.147	0.354
East North Central (ENCEN)	0.174	0.379
East South Central (ESCEN)	0.096	0.294
West North Central (WNCEN)	0.156	0.363
West South Central (WSCEN)	0.161	0.368
Mountain (MTN)	0.072	0.258
Pacific (PAC)		
excluded from regression	0.126	0.332
Government owned hospital (GOVT)	0.296	0.456
Proprietary hospital (PROFIT)	0.147	0.355
Voluntary hospital (NONPROF)		
excluded from regression	0.557	0.497

NOTE: AHA is American Hospital Association.

SOURCE: 1984 Medicare cost reports and 1985 AHA Hospital Survey.

highly skewed. Alternative models were tested with the objective of making the distribution of cost per discharge symmetric (Box and Cox, 1962). Upon examination, the distribution of the natural logarithm of average costs is approximately normally distributed. Moreover, assuming a loglinear (multiplicative) model generates studentized residuals that are also approximately normally distributed.¹ In this loglinear model, using logarithms of the continuous independent variables leads to a better fit and to a more natural interpretation of the results.

If the Medicare case-mix index is compressed, its use in the loglinear cost function would generate a

¹Use of linear splines suggests that the estimated elasticities of cost per discharge with respect to case mix are significantly less than one in hospitals with high case-mix values. Factors accounting for these differences are not clear.

biased estimate of the coefficient on case mix. Yet, if the degree of compression in the current case-mix index could be estimated, the existing case-mix values could be adjusted. For instance, the loglinear cost function yields the following estimating equation (for D_j = cost per discharge at hospital j)

$$\ln(D_j) = \ln(a) + B_0 \ln(C_j') + B_1 \ln(X_1) + \dots + B_n \ln(X_n), \quad (4)$$

where case mix should be defined so that the estimated coefficient of $\ln(C_j')$, B_0 , equals 1. Yet, if the cost function described in equation (4) is estimated using the compressed case mix C_j instead of the true case mix C_j' , B_0 will not be 1. Instead, it will be a good estimate of the degree of compression. Note that if we define S_j by $C_j = 1 + S_j$, then we can rewrite equation (3) as follows:

$$C_j' = 1 + kS_j, \quad (5)$$

For S_j small, and k near 1, we have an excellent approximation

$$(1 + S_j)^k \sim 1 + kS_j. \quad (6)$$

Taking logs, we obtain

$$\ln(C_j') \sim k \ln(C_j). \quad (7)$$

Substitution of equation (7) into equation (4) with $B_0 = 1$ gives

$$\ln(D_j) = \ln(a) + k \ln(C_j) + \sum B_j \ln(X_j). \quad (8)$$

Thus, the coefficient of $\ln(C_j)$ is a good estimate of the degree of compression. If the current charge-based weights are still compressed, the estimated coefficient of the compressed case-mix index will exceed 1.

Results

The loglinear cost functions relating Medicare operating cost per discharge to the exogenous variables described previously were estimated, and their results are displayed in Table 3. The coefficient estimates of log DRG are generally of the expected direction, and they provide some evidence that the DRG relative prices remain compressed.

The two coefficient estimates—those for the case-mix index and the ratio of interns and residents per bed—are sensitive to whether or not hospital bed size is included in the estimating equation (compare Models 2 and 3). When including only the HCFA payment variables in the cost function, the coefficient estimate on case mix is 1.24, significantly greater than 1.² Adding dummy variables for census region or a

²Data on the percent of Medicare days that were covered under the Federal Supplemental Security Income program were not available. Hence, we were unable to include the exact measure of the disproportionate share adjustment in the model. Inclusion of percent of Medicaid days in Model 1, however, was insignificant and did not alter other estimated coefficients.

Table 3
Estimated Medicare operating cost per discharge coefficient, estimated coefficients, and standard errors

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	*7.74 (0.006)	*7.74 (0.013)	*7.51 (0.025)	*7.73 (0.014)	*7.46 (0.023)
Log (DRG)	†1.24 (0.030)	†1.23 (0.030)	*1.02 (0.036)	†1.23 (0.031)	*1.04 (0.036)
Log (1 + INT)	*0.73 (0.057)	*0.72 (0.057)	*0.62 (0.058)	*0.71 (0.059)	*0.63 (0.059)
Log (WAGE)	†1.16 (0.033)	†1.13 (0.044)	*1.09 (0.044)	†1.13 (0.045)	*0.92 (0.048)
NONMET	*-0.05 (0.010)	*-0.06 (0.011)	-0.03 (0.011)	*-0.06 (0.011)	
NENG		*0.05 (0.023)	0.02 (0.023)	*0.05 (0.023)	0.03 (0.023)
MIDAT		*-0.04 (0.017)	*-0.08 (0.017)	-0.04 (0.017)	*-0.08 (0.017)
SOATL		-0.001 (0.014)	*-0.05 (0.015)	*-0.00 (0.014)	*-0.06 (0.015)
ENCEN		*0.06 (0.012)	0.01 (0.013)	*0.06 (0.012)	0.02 (0.013)
ESCEN		** -0.03 (0.017)	*-0.08 (0.015)	** -0.03 (0.017)	*-0.09 (0.017)
WNCEN		0.014 (0.015)	-0.004 (0.015)	0.02 (0.015)	-0.01 (0.015)
WSCEN		*0.004 (0.014)	** -0.03 (0.014)	0.005 (0.015)	*-0.04 (0.015)
MTN		0.011 (0.016)	0.001 (0.015)	0.013 (0.016)	0.004 (0.016)
Log (BEDS)			*0.06 (0.005)		*0.05 (0.005)
Log (1 + CAID)				0.04 (0.055)	*0.14 (0.054)
GOVT					*-0.02 (0.007)
PROFIT					*0.06 (0.010)
LMA					*0.09 (0.015)
NMA					*0.03 (0.012)
SMA					**0.02 (0.013)
Adjusted R-Square	0.659	0.664	0.672	0.664	0.679

*Significantly different from zero, $P \leq 0.05$ (two-tailed test).

**Significantly different from zero, $P \leq 0.10$ (two-tailed test).

†Significantly different from one, $P \leq 0.05$.

NOTES: Pacific omitted from regression. 102 of the 5,015 hospitals reported in Table 2 had missing AHA data. These hospitals were excluded from all regressions in Tables 3 and 5. Exclusion of these observations resulted in trivial differences in estimated parameters (Table 6). Definitions of variables are given in Table 2. Numbers in parentheses are standard errors. AHA is American Hospital Association.

SOURCE: 1984 Medicare cost reports and 1985 AHA Hospital Survey.

variable for the percent Medicaid has no significant effect on the case-mix coefficient. However, adding the bed-size variable reduces the coefficient on case mix to 1.02, which is not statistically different from 1. The estimated coefficient on the ratio of interns and residents per bed follows a similar pattern, ranging from 0.71 to 0.73 when bed size is omitted and dropping to 0.62 to 0.63 when bed size is included.

Our estimate of the degree of compression in the Medicare case-mix index is clearly quite sensitive to which variables are included in the cost function.

Although the more complete models provide the best estimate of irreducible compression, the more parsimonious models, Model 1 and Model 2, provide better estimates of the operational compression in the index under the Medicare prospective payment system (PPS) with national and regional rates, respectively. Operational compression includes the three sources of compression discussed previously as well as the influence of relevant variables omitted from the model. Model 4 represents PPS under regional rates

with an adjustment for the proportion of Medicaid days.

In comparing our results with the earlier work by Pettengill and Vertrees (1982) and Lave (1985b), in which the old cost-based case-mix index and older hospital data were used, we find that our Models 3 and 5, with specifications similar to theirs, yield similar results. In Model 5, a 1-percent increase in the Medicare case-mix index is associated with approximately a 1.04-percent increase in Medicare cost per discharge. The estimated coefficient on case mix is not significantly different from 1 when using a standard 95-percent confidence interval. Pettengill and Vertrees (1982) found an estimated coefficient of 1.08 using 1979 data, which was about 1.8 standard deviations above 1. However, a subsequent analysis by Lave (1985b) appears to reduce the compression in the DRG case-mix values, but some compression remains. The similarity of our findings suggests that whatever compression actually exists was not affected by more accurate coding of the DRG categories over time.

Looking at Model 5, the proportion of Medicaid days is also positively related to Medicare cost per discharge, although the implied effect is relatively small. For instance, the relationship suggests that, if the proportion of Medicaid discharges goes from 0 to 10 percent, then Medicare cost per discharge increases by less than 2 percent.

If the wage rate is used to adjust part of the DRG payment, the coefficient should be less than 1. However, the estimated coefficient on wage exceeds 1 in three of the five models reported. The coefficient on wage is significantly lower, however, when city size is included in the regression equation. Perhaps the reason that the coefficient on wage is greater than 1 in the other models is that it serves as a proxy for city size, which increases costs.

Hospital location also influences Medicare cost per discharge. Even after adjusting for differences in wages, hospitals in large urban areas have higher costs per discharge than hospitals in rural areas. Other factors held constant, costs per discharge for hospitals located in large metropolitan areas are approximately 9.4 percent higher than costs in rural areas.³ Cost was per discharge in other urban areas was approximately 2 to 3 percent higher than that in rural areas.

Medicare cost per discharge also varies by region. Much of the observed regional variation in cost may be attributed to large variations in length of stay, ancillary utilization, and other factors influenced by physicians (Rothberg, 1982; Mitchell, 1985). These cost differentials may also reflect systematic differences in hospital efficiency which differ by region. Hospitals located in the South Atlantic, Middle Atlantic, and East South Central experienced costs per Medicare discharge of 6.8 and 9 percent lower, respectively, than those in the Pacific.

³Correct interpretation of dummy variables in semilogarithmic equations requires an adjustment of the estimated coefficients. For a discussion, see Kennedy (1980).

Finally, hospital ownership appears related to Medicare cost per discharge. Specifically, cost per discharge for government hospitals appears approximately 2 percent lower than that for nonprofit hospitals. Further, Medicare cost per discharge in for-profit institutions is approximately 6 percent higher than that in nonprofit hospitals. Because reimbursement is the same, this implies lower profits for proprietaries.

Discussion

Evidence for compression

From equation (8), we see that the coefficient on the log of the case-mix index gives an estimate of compression. The results of our most complete model (Model 5, Table 3) suggest that the standard deviation of the true case-weight distribution is only 4 percent higher than that of the distribution of case weights based on charges. However, with the model most closely in line with the PPS payment formula, the standard deviation of the true case-weight distribution is 24 percent higher than that of the charge-based case weights.

Sample values of the current compressed charge-based, case-mix values to values under decompressed case-mix index, adjusted by applying equation (3), are given in Table 4. As implied by equation (3), the percent error in the current charge-based, case-mix index values increases for values further from 1. Payments to a hospital would, of course, change in direct proportion to the changes in the hospital's case-mix index.

Table 4
Comparison of sample Medicare charge-based, case-mix index to adjusted case-mix index

Charge-based case-mix index	Decompressed case-mix index	
	Based on Model 1	Based on Model 5
0.750	0.690	0.740
0.900	0.876	0.896
1.000	1.000	1.000
1.100	1.124	1.104
1.250	1.310	1.260
1.500	1.620	1.520

Is decompressed case-mix index valid?

With our estimates of compression, we can decompress the current case-mix index using equation (3), replace the compressed case-mix index by the decompressed index in the cost function, and reestimate the equation. If the adjusted relative weights provide a more accurate measure of the true case-mix index, the estimated coefficient in the cost function should decrease and be close to 1. The results of this adjustment are displayed in Table 5.

As expected, the estimated coefficient on the case-mix index is much closer to 1 when the decompressed

Table 5

estimated Medicare operating cost per discharge with adjusted decompressed Medicare case-mix estimated coefficients and standard errors for 4,913 of the hospitals in the 1984 AHA Hospital Survey

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Compression (k)	1.24	1.23	1.02	1.23	1.04
Constant	*7.74 (0.006)	*7.74 (0.013)	*7.51 (0.024)	*7.73 (0.14)	*7.46 (0.023)
Log (DRG)	*1.00 (0.025)	*1.00 (0.025)	*1.00 (0.035)	*1.00 (0.025)	*1.00 (0.034)
Log (1 + INT)	*0.74 (0.057)	*0.74 (0.057)	*0.62 (0.058)	*0.74 (0.059)	*0.63 (0.059)
Log (WAGE)	*1.16 (0.033)	*1.13 (0.045)	*1.09 (0.044)	*1.13 (0.045)	*0.92 (0.048)
NONMET	*-0.05 (0.010)	-0.06 (0.011)	*-0.03 (0.011)	*-0.06 (0.011)	
NENG		*0.05 (0.023)	0.02 (0.023)	*0.05 (0.023)	0.03 (0.023)
MIDAT		*-0.05 (0.017)	*-0.08 (0.017)	*-0.04 (0.017)	*-0.08 (0.017)
SOATL		-0.002 (0.014)	*-0.05 (0.015)	-0.001 (0.014)	*-0.06 (0.015)
ENCEN		*0.06 (0.013)	0.02 (0.013)	*0.06 (0.012)	0.02 (0.015)
ESCEN		** -0.03 (0.017)	*-0.08 (0.017)	** -0.03 (0.017)	*-0.09 (0.017)
WNCEN		0.013 (0.015)	-0.004 (0.015)	0.015 (0.015)	-0.01 (0.015)
WSCEN		0.003 (0.015)	** -0.03 (0.014)	0.004 (0.014)	-0.04 (0.015)
MTN		0.01 (0.016)	-0.001 (0.016)	0.01 (0.016)	0.003 (0.016)
Log (BEDS)			*0.06 (0.005)		*0.05 (0.005)
Log (1 + CAID)				0.03 (0.055)	*0.14 (0.054)
GOVT					*-0.02 (0.007)
PROFIT					*0.06 (0.010)
LMA					*0.09 (0.014)
NMA					*0.03 (0.012)
SMA					*0.02 (0.012)
R-Square	0.658	0.664	0.672	0.663	0.679

*Significantly different from zero, $P \leq 0.05$ (two-tailed test).

**Significantly different from zero, $P \leq 0.10$ (two-tailed test).

NOTES: Pacific omitted from regression. Definitions of variables are given in Table 2. Numbers in parentheses are standard errors. AHA is American Hospital Association.

SOURCE: 1984 Medicare cost reports and 1985 AHA Hospital Survey.

index is used. Thus, the corrected case-mix index corrects for both operational and irreducible compression. A second finding is that using the decompressed index in place of the compressed index has no noticeable effect on the coefficients of the other variables in the cost function. This null result is important because, in theory, at least, misspecification of the case-mix variable in the cost equation could have biased the coefficient estimates for other variables that are strongly correlated with case mix. The null result means that the results presented by Pettengill and Vertrees (1982), which were initially used in developing the DRG payment rates, were probably not greatly biased by compression in the case-mix index.

The results in Tables 3 and 5 underscore the robustness of the compressed case-mix index from a

statistical point of view. Use of the compressed index in any given cost equation does not alter the explanatory power or the coefficient estimates for other variables when compared with the same model using a decompressed index.

Payment implications

This does not mean that the choice of a compressed or decompressed index is unimportant from a practical or political point of view, however. The choice between a compressed or a decompressed index will affect which hospitals find themselves overpaid or underpaid. Moreover, payment levels will also depend on the specific decompressed case-mix index selected (Table 5). Recognition of operational compression as well as the implicit recognition of those determinants

of cost (e.g., those positively correlated with case mix and costs) not currently used to determine payment rates would eliminate irreducible compression. Under a budget neutral system, any adjustments to the case-mix values will affect the distribution of Medicare payments across hospitals. If the current case-mix values were decompressed, hospitals with an above average case mix would receive increased payments, and payments to hospitals with below average case mix will fall. Moreover, decompression would also adjust the structure of relative prices, potentially influencing the profitability of certain DRG's and, as a result, the pattern of hospital admissions.

The sensitivity of our results to what is included in the model raises an important question about the nature of what has been termed "compression." Inclusion of hospital bed size as an explanatory variable greatly reduces the compression in the relative prices. To what extent do bed size and compression measure the same phenomenon? The results of four regressions run on variations of our most parsimonious model are shown in Table 6. These regressions have been run with compressed and decompressed case-mix indexes and with and without bed size included as an independent variable. The results show that whatever is contributing to operational compression in the case-mix index is nearly completely resolved by leaving the case-mix index compressed and instead adding bed size to the equation. Furthermore, adding bed size to the model using the decompressed index reduces the coefficient

on the case-mix index to 1, further confirming that bed size and compression are two different ways of measuring the same thing. What that thing is, however, is not so clear.

Limitations in the methods

The methods and data we have used to look for compression in the case-mix index have limitations that need to be borne in mind when interpreting our results. For one thing, the cost per discharge variable employed here was based on a methodology similar to that used in Pettengill and Vertrees (1982). This means that our cost data will not reveal compression in the index because of the assumption that routine and special care per diem costs are the same for all DRG's. A second limitation is our assumption that compression is uniform. A more sophisticated model, allowing the degree of compression to vary with case weight, might give somewhat different results.

Conclusions

Compression is important because after hospital costs are standardized, Medicare's relative price schedule determines the marginal revenues hospitals receive for providing services for each of 471 different DRG's. This payment formula naturally attracts significant attention toward the structure of the relative prices. The analysis presented indicates that compression in the Medicare relative prices results in lower payment rates to high case-mix hospitals and higher payments to low case-mix hospitals. The same distortion also changes the profitability of individual DRG's as high-cost DRG's are priced too low and low-cost DRG's are priced too high.

Our analyses suggest that the degree of compression in the case-mix index is largely determined by which auxiliary variables, in addition to case mix, are used to explain costs. The switch to charge-based weights and the documented changes in coding between 1981 and 1984 (Carter and Ginsburg, 1985) have had minor effects in reducing the level of compression in hospital cost equations modeled after Pettengill and Vertrees (1982), (Cotterill, Bobula, and Connerton, 1986). If coding problems or the use of the old cost-based weights had been the primary cause of compression, then our analysis would have shown less compression than the earlier work (Pettengill and Vertrees, 1982; Lave, 1985b). Because more complete coding and charge-based weights did not eliminate compression, the cause must lie elsewhere. Even more disturbing, however, is the finding that the cost equation modeled closely on the prospective payment variables shows operational compression of 24 percent in the case-mix index.

The analysis provides estimates for two different types of case-mix compression, irreducible and operational. Irreducible compression results from the three problems noted earlier—the assumption of constant per diem costs across DRG's, cross-subsidization, and classification errors. Our

Table 6

Estimated Medicare operating cost per discharge with and without bed size for compressed and decompressed case-mix index estimated coefficients and standard error for 5,015 of the hospitals in the 1984 AHA Hospital Survey

Variable	Without bed size		With bed size	
	Com-pressed	Decom-pressed	Com-pressed	Decom-pressed
Decompression factor (k)	1.00	1.24	1.00	1.07
Constant	*7.74 (0.006)	*7.74 (0.006)	*7.50 (0.022)	*7.51 (0.23)
Log (compressed DRG)	*1.24 (0.031)		*1.07 (0.034)	
Log (decompressed DRG)		*0.99 (0.027)		*1.00 (0.034)
Log (1 + INT)	*0.73 (0.058)	*0.74 (0.058)	*0.60 (0.058)	*0.60 (0.058)
Log (WAGE)	*1.17 (0.033)	*1.17 (0.033)	*1.19 (0.032)	*1.19 (0.032)
NONMET	*-0.05 (0.010)	*-0.05 (0.010)	-0.011 (0.010)	-0.012 (0.010)
Log (BEDS)			*0.05 (0.005)	*0.05 (0.005)
R-Square	0.652	0.652	0.660	0.660

*Significantly different from zero, $P \leq 0.05$ (two-tailed test).

NOTES: Definitions of variables are given in Table 2. Numbers in parentheses are standard errors. AHA is American Hospital Association.

SOURCE: 1984 Medicare cost reports and 1985 AHA Hospital Survey.

results indicate the extent of such compression noted by Pettengill and Vertrees (1982) is likely quite small. Operational compression in the case weights reflects variables omitted from hospital cost functions as well as the problems noted before. It is clear that what looks like compression in the more parsimonious models may be related to any number of underlying differences between high and low case-mix hospitals not currently recognized by Medicare in the prospective payment system. Thus, the decision to use a decompressed case-mix index depends on the source of compression to be eliminated. Isolating the source and policy relevance of the differences in irreducible and operational compression is an important area for future research. If these differences are the result, in part, of variations in hospital efficiency, then eliminating operational compression may not be desirable.

Acknowledgments

We are indebted to Joseph P. Newhouse, Gus Haggstrom, Judy Lave, and three anonymous referees for their helpful comments; to Chris Witsberger for her invaluable assistance in preparing the analytic files; and to Martha Cooper for secretarial assistance. We would also like to thank Stephen Jencks of the Health Care Financing Administration's Office of Research for helpful comments on an earlier draft.

References

- Becker, E., and Sloan, F.: Hospital ownership and performance. *Economic Inquiry* 23(1):21-36, Jan. 1985.
- Belsley, D., Kuh, E., and Welsch, R.: *Regression Diagnostics: Identifying Influential Data and Source of Collinearity*. New York. Wiley Press, 1980.
- Box, G., and Cox, D.: An analysis of transformations. *Journal of Royal Statistical Society, Series B*, 1962.
- Carter, G., and Ginsburg, P.: *The Medicare Case-Mix Index Increase: Medical Practice Changes, Aging, and DRG Group*. The RAND Corporation, R-3292-HCFA, June 1985.
- Caves, D., and Christensen, L.: Flexible cost functions for multi-product firms. *Review of Economics and Statistics* 62(1):477-481, Oct. 1980.
- Cotterill, P., Bobula, J., and Connerton, R.: Comparison of alternative relative weights for diagnosis-related groups. *Health Care Financing Review*. Vol. 7, No. 3. HCFA Pub. No. 03222. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Spring 1986.
- Cowing, T., and Holtmann, A.: The multiproduct short-run hospital cost function: Empirical evidence and policy cross section data. *Southern Economic Journal* 49(1):637-653, Jan. 1983.
- DeAlissi, L.: The economics of property rights: A review of the evidence. *American Economic Review* 73(1):64-81, Mar. 1983.
- Federal Register*: Vol. 48, No. 171, pp. 39871-5. Office of the Federal Register, National Archives and Records Administration. Washington. U.S. Government Printing Office, Sept. 1, 1983.
- Harris, J.: Pricing rules for hospitals. *Bell Journal of Economics* 10(1):224-243, Spring 1979.
- Institute of Medicine: *Reliability of Medicare Hospital Discharge Records*. National Academy of Sciences. NTIS PB 281680, Nov. 1977.
- Kennedy, P.: Estimation of correctly interpreted dummy variables in semilogarithmic equations. *American Economic Review* 71(4):801, Sept. 1981.
- Lave, J.: Is compression occurring in the DRG prices? *Inquiry* 22(2):142-147, Summer 1985a.
- Lave, J.: *The Medicare Adjustment for the Indirect Costs of Medical Education: Historical Development and Current Status*. Association of American Medical Colleges, Washington, D.C., Jan. 1985b.
- Mitchell, J.: Physician DRGs. *The New England Journal of Medicine* Vol. 313, Sept. 1985.
- Newhouse, J., Cretin, S., and Witsberger, C.: *Predicting Hospital Accounting Costs Using the Synthetic Cost Method: A Validation Study*. The RAND Corporation, May 1987.
- Office of Technology Assessment, U.S. Congress: Diagnosis related groups and the Medicare program. *Implications for Medical Technology*. Washington. U.S. Government Printing Office, 1983.
- Pettengill, J., and Vertrees, J.: Reliability and validity in hospital case-mix measurement. *Health Care Financing Review*. Vol. 4, No. 2. HCFA Pub. No. 03149. Office of Research, Demonstrations, and Statistics, Health Care Financing Administration. Washington. U.S. Government Printing Office, Winter 1982.
- Prospective Payment Assessment Commission: *Technical Appendixes to the Report and Recommendations to the Secretary*. U.S. Department of Health and Human Services, Apr. 1, 1985.
- Rothberg, D., ed.: *Regional Variations in Hospital Use: Geographic and Temporal Patterns in the U.S.* Boston. D.C. Health, 1982.
- Sloan, F., Feldman, R., and Steinwald, B.: The effects of teaching on hospitals costs. *Journal of Health Economics*, Mar. 1983.
- Wagner, P., and Draper, E.: Acute physiology and chronic health evaluation (APACHE II) and Medicare reimbursement. *Health Care Financing Review*. 1984 Annual Supplement. HCFA Pub. No. 03194. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Dec. 1984.