

Acute physiology and chronic health evaluation (APACHE II) and Medicare reimbursement

by Douglas P. Wagner and Elizabeth A. Draper

This article describes the potential for the acute physiology score (APS) of acute physiology and chronic health evaluation (APACHE) II, to be used as a severity adjustment to diagnosis-related groups (DRG's) or other diagnostic classifications. The APS is defined by a relative value scale applied to 12 objective physiologic variables routinely measured on most hospitalized patients shortly after hospital admission.

For intensive care patients, APS at admission is strongly related to subsequent resource costs of intensive care for 5,790 consecutive admissions to 13 large hospitals, across and within diagnoses. The APS could also be used to evaluate quality of care, medical technology, and the response to changing financial incentives.

Introduction

As prospective reimbursement of hospitals based on diagnosis-related groups (DRG's) is implemented, there remains concern about interhospital variations of cost-increasing severity of illness within DRG's.

This article describes the potential for APACHE II, a simplification of APACHE (acute physiology and chronic health evaluation) to be used as a reimbursement supplement to DRG's or other diagnostic classifications. Systematic collection of 12 objective physiological measures for all patients at hospital admission probably could improve the resource homogeneity and clinical acceptability of revised DRG categories. Such information could add more clinical precision to discussions among physicians about appropriate use of resources for specific patients. APACHE II could also substantially improve the government's ability to monitor the response to the changed financial incentives, to evaluate new technology, and to evaluate quality of care across hospitals.

The next section of this article briefly discusses the evidence suggesting a need for severity measurement. Subsequent sections describe the rationale underlying APACHE II and prior results, the measurement of APACHE II, the relationship between APACHE II and resource costs, further research needs for integration with DRG's, potential advantages of using APACHE II, and conclusions.

Need for severity measurement in reimbursement

There is considerable belief that, within DRG categories, patients are not randomly allocated across hospitals. Many physicians and health services researchers suspect that within a given medical diagnosis or surgical procedure, the patients requiring

more complex management are more often hospitalized at larger, teaching hospitals with a wider variety of specialized services (Garber *et al.*, 1984). This could occur because 1) attending physicians refer patients with more complicated illnesses to specialists at tertiary care settings, 2) physicians have admitting privileges to more than one hospital and admit the more severely ill patients to the larger hospital, or 3) patients who believe they have a difficult health problem select physicians associated with larger hospitals. In addition, patients who desire more aggressive care might choose a physician at a larger teaching hospital for their attending physician regardless of the severity of illness.

A detailed examination of the interhospital validity of DRG's for prospective reimbursement was conducted by Pettengill and Vertrees (1982). They constructed a DRG-based case-mix index using a 20 percent sample of Medicare inpatient bills for each of the 5,071 hospitals with 50 or more sample Medicare discharges in 1979. Adjusting for capital costs, they found the Medicare operating cost per case to be significantly and substantially related to the hospital DRG case-mix index, hospital size, number of interns and residents per bed, wage rates, and urban size. The elasticity of adjusted costs with respect to the hospital DRG case-mix index was 1.08, not significantly different from 1.0. This supports the hypothesis that DRG's are an adequate measure of interhospital variations in case mix for reimbursement purposes.

There is, however, some evidence that there are substantial interhospital variations in severity and cost per case not captured by the DRG case-mix index (Horn, 1983). On a substantially larger data base, the multivariate analysis of Pettengill and Vertrees (1982) establishes that several variables other than DRG case weight are significantly and substantially associated with interhospital variations in Medicare costs. One interpretation of their work is that hospital size, teaching intensity, and urban size are proxies, in part, for severity of illness. These independent variables are highly correlated with average cost and DRG case weight, but are not all reimbursed under prospective payment.

This study was supported by the National Center for Health Services Research, U.S. Department of Health and Human Services, Grant HS 04857. Additional support came from the Robert Wood Johnson Foundation, Grant No. 8498.

Reprint requests: Douglas P. Wagner, ICU Research, 2300 K Street, N.W., Washington, D.C. 20037

The issues are best illustrated by comparing the Medicare prospective payment reimbursement with the average cost predicted by the equation in the Pettengill and Vertrees analysis (1982). The critical variable included in the regression analysis but excluded from the reimbursement formula is hospital bed size. The elasticity of cost with respect to hospital size is substantial and extraordinarily significant. It implies that a typical 500-bed hospital has a Medicare cost per case which is about 21 percent more expensive than would be predicted based on DRG case mix, location, and teaching intensity alone. Because large hospitals tend to have high hospital DRG case-mix weights, the above differential is compounded by the difference between the estimated hospital DRG case-mix coefficient of 1.08 and the theoretically fair value of 1.0. For a hospital with a large case-mix index, for example 20 percent larger than average, the difference between 1.00 and 1.08 causes a 2-percent discrepancy between cost and reimbursement. Hospitals below the mean in bed size or DRG case-mix index are correspondingly rewarded with reimbursement greater than projected cost.

Most large, tertiary care, teaching hospitals are located in large urban areas. In the Pettengill and Vertrees analysis, location in a large urban area also implies a 15 percent increase in cost per case. Normally an econometrician would expect an urbanization variable to capture regional price variations. The equation, however, also included a separate price variable, wage rates, which was extraordinarily significant with a coefficient of 1.00. If one believes a priori that regional wage and price variations are passed directly through to cost variations (the production of hospital care is homogeneous to degree 1 in input prices) then the estimated wage elasticity of 1.00 strongly confirms that belief. The precision of this wage coefficient suggests that urbanization does not capture price differences. Therefore, urban size may also imperfectly reflect severity of illness differences.

The potential under-reimbursement for the cost-increasing severely ill patients in large teaching hospitals is partially corrected by the doubling of the reimbursement for the indirect cost of graduate medical education and the use of an urban/rural price differential under the prospective payment system.

It is possible that larger hospitals are simply less efficient producers of hospital care, despite the potential for economies of scale and the long-held belief that volume and experience are correlated with effectiveness and efficiency. Previous reimbursement policies provided little incentive to be cost effective. However, two alternative interpretations appear more attractive than the former.

One hypothesis is that the hospital accounting data, even the Medicare provider analysis and review (MEDPAR) file and Medicare cost report data, are simply too inaccurate to evaluate the questions. Historically, at least some hospitals' accounting systems have been designed to maximize revenue from cost-paying third parties such as Medicare and most Blue Cross plans. This results in inflated prices for

frequently used items and blurs the relationship between accounting costs and economic costs of production (Finkler, 1982). Also, larger hospitals do maintain the capacity to do a number of rarely used and difficult procedures and diagnostic tests. These costs are also rolled into departmental accounting and inflate the prices of frequently used items and services (Williams et al., 1982; Wagner, Wineland, and Knaus, 1983).

The other hypothesis is that hospital size, urban size, and teaching intensity are partial proxies for cost-increasing severity of illness which is not captured by DRG categories. One method to explore the latter hypothesis is the subject of the rest of this article.

APACHE

Development of the original APACHE (acute physiology and chronic health evaluation) severity-of-illness classification system began in 1978 with the specific goal of developing a measure for use in describing groups of intensive care unit (ICU) patients and evaluating their care. ICU's receive patients with a wide variety of diagnoses and severity of illness, and it is difficult for one ICU physician to precisely describe his case mix to another. Diagnoses are necessary but not sufficient.

Because APACHE was designed for the evaluation of efficacy of medical treatment, the timing, quality, and type of data collected have been different than in research principally oriented toward hospital reimbursement questions. The most important difference is that all of the severity and diagnostic data have been collected early in the course of each patient's hospital stay, within 24 hours of ICU admission, rather than after hospital discharge. In medical research terminology, this has been a prospective observational study, not retrospective chart review.

The underlying philosophy of APACHE is that the wide variety of physiologic measurements routinely obtained on ICU patients contain precise information on the patient's acute severity of illness. Therefore, the original APACHE consisted of an acute physiology score (APS) based on 34 physiologic variables and a chronic health assessment (Knaus, Zimmerman *et al.*, 1981). The latter was a separate 4-category scale derived from items previously used to assess chronic health by the Health Interview Survey, the Rand Health Insurance Study, and the New York Heart Index.

Consensus of a group of experienced ICU clinicians was used to select the 34 APS variables and to specify how to weight derangements in each. Several of the 34 variables are measured only on patients with specific diagnoses or symptoms. We therefore made the important assumption that variables unmeasured in the ICU setting are unlikely to be seriously deranged and can be assumed to be normal. This assumption appears to be reasonably accurate in most hospital ICU's that we have sampled. This is particularly true for measurement during the first day in the ICU when

a wide variety of physiologic variables are routinely repeatedly measured.

The translation of the 34 variables into APS weights is illustrated for respiratory rate below. Note that there is a wide range of normality which receives a weight of zero and that the weights increase in a nonsymmetric and nonlinear manner as the patient's breaths per minute varies from the normal range in either direction.

APS weights for respiratory rate
(Breaths per minute)

Respiratory rate	5<	6-9	10-11	12-24	25-34	35-49	50<
APS weight	4	2	1	0	1	3	4

The patient's APS score is the sum of the weights for the most deranged value of each of the 34 variables measured within the first 24 hours of ICU admission.

Initial results with APACHE on the first 600 consecutive ICU admissions at George Washington University Medical Center (GW) were quite promising. The patients were widely distributed across the APS score, from 0 to the high 50's, with a bell-shaped distribution. The APS was highly significantly related to whether the patient was dead or alive at hospital discharge (Knaus, Zimmerman et al., 1981). It was expected that APS would also be significantly related to the resource costs of treatment in the ICU. The more severely ill patients do receive more complex therapy and it is likely to take the patient longer to recover. This hypothesis was also strongly supported by the data.

Subsequent research demonstrated that APACHE could be measured in a community hospital with equal precision and predictive validity, but that the community hospital had far different patients in its ICU than did GW (Draper et al, 1981). Most of the community hospital ICU patients were there to be observed closely, not aggressively treated. Despite having similar medical diagnoses, their average severity of illness was only 7 APS points compared to 16 at GW Hospital. The APS of APACHE was by far the most significant variable in explaining variation in survival and resource cost of care. It accounted for more than 50 percent of the explained variation in each dependent variable, and its regression coefficient was affected very little by the inclusion or deletion of a number of diagnostic and other clinical variables (Draper et al., 1981).

The APS of APACHE is also sensitive at the lower range of severity of illness. It is capable of accurately identifying which ICU patients who were admitted for monitoring were at low risk of ever needing aggressive and unique ICU therapy (Wagner, Knaus, Draper, and Zimmerman, 1983; Knaus, Draper, and Wagner, 1983).

Further research with this measure revealed that it could be measured reliably in a number of hospitals, and that severity of illness in the first 24 hours of ICU admission could accurately predict variations

from 7 to 30 percent in hospital death rates. (Knaus, Draper *et al.*, 1982). The measure proved quite useful in comparing ICU case mix and medical practice differences between the United States and France (Knaus, Le Gall *et al.*, 1982). It was also demonstrated that the APS was significantly associated with outcome with approximately the same magnitude within a number of specific cardiovascular, neurologic, respiratory, and gastrointestinal diagnoses (Wagner, Knaus, and Draper, 1983).

APACHE II

In 1982 prospective data collection was begun of samples of 200 to 500 consecutive ICU admissions to 12 other hospitals across the country. One of the major objectives of this effort was to obtain sufficient data to do a rigorous examination of whether the APS measure could be substantially simplified without loss of precision.

The result of the simplification effort is APACHE II, which is based on 12 of the most commonly measured physiologic measures included in the original APACHE system (Knaus *et al.*, submitted for publication, 1984a). The 12 variables were selected based on clinical judgement as to validity and specificity of the measure, breadth of vital organ system coverage, and objectivity, reliability and frequency of measurement. The 12 variables include vital signs (heart rate, mean blood pressure, respiratory rate, temperature, and Glasgow Coma Score), variables derived from routine venous blood tests (hematocrit and white blood cell count, serum potassium, serum sodium, and serum creatinine), and 2 variables derived from arterial blood gas tests (serum pH and PaO₂). Full details are reported in Figure 1.

Most of these 12 variables are routinely measured on most hospital patients shortly after hospital admission. The exceptions are serum creatinine and the blood gas values. Some hospitals substitute the more sensitive but less specific serum BUN for serum creatinine in the SMA-6, a standardized automated blood test that produces measures of 6 blood serum components. If so, the serum creatinine is usually included in the SMA-12, a slightly more complex standardized automated blood test. For patients in whom oxygenation is normal and blood gases not measured, an HC03 from the SMA-6 can be used in lieu of the serum pH in the calculation of an APS score.

Each of these 12 variables is translated into weights using the original APACHE relative value scale with slight modifications. Thus the core of APACHE II remains the systematic application of clinical judgement about the relative importance of derangement in the physiologic measures. APACHE II also assigns weights to increased age and severe chronic disease and integrates them into a single integer score. This score is strongly related to hospital survival among the 5,815 ICU patients in the data base. For hospital reimbursement purposes, however, it would probably

Figure 1

The acute physiology and chronic health evaluation (APACHE) II severity of disease classification system

PHYSIOLOGIC VARIABLE	HIGH ABNORMAL RANGE					LOW ABNORMAL RANGE			
	+4	+3	+2	+1	0	+1	+2	+3	+4
TEMPERATURE — rectal (°C)	○ ≥ 41°	○ 39°-40.9°		○ 38.5°-38.9°	○ 36°-38.4°	○ 34°-35.9°	○ 32°-33.9°	○ 30°-31.9°	○ ≤ 29.9°
MEAN ARTERIAL PRESSURE — mm Hg	○ ≥ 160	○ 130-159	○ 110-129		○ 70-109		○ 50-69		○ ≤ 49
HEART RATE (ventricular response)	○ ≥ 180	○ 140-179	○ 110-139		○ 70-109		○ 55-69	○ 40-54	○ ≤ 39
RESPIRATORY RATE — (non-ventilated or ventilated)	○ ≥ 50	○ 35-49		○ 25-34	○ 12-24	○ 10-11	○ 6-9		○ ≤ 5
OXYGENATION: A-aDO ₂ or PaO ₂ (mm Hg) a. FiO ₂ ≥ 0.5 record A-aDO ₂ b. FiO ₂ < 0.5 record only PaO ₂	○ ≥ 500	○ 350-499	○ 200-349		○ ≤ 200				
					○ PO ₂ > 70	○ PO ₂ 61-70		○ PO ₂ 55-60	○ PO ₂ < 55
ARTERIAL pH	○ ≥ 7.7	○ 7.6-7.69		○ 7.5-7.59	○ 7.33-7.49		○ 7.25-7.32	○ 7.15-7.24	○ ≤ 7.15
SERUM SODIUM (mMol/L)	○ ≥ 180	○ 160-179	○ 155-159	○ 150-154	○ 130-149		○ 120-129	○ 111-119	○ ≤ 110
SERUM POTASSIUM (mMol/L)	○ ≥ 7	○ 6-6.9		○ 5.5-5.9	○ 3.5-5.4	○ 3-3.4	○ 2.5-2.9		○ ≤ 2.5
SERUM CREATININE (mg/100 ml) (Double point score for acute renal failure)	○ ≥ 3.5	○ 2-3.4	○ 1.5-1.9		○ 0.6-1.4		○ ≤ 0.6		
HEMATOCRIT (%)	○ ≥ 60		○ 50-59.9	○ 46-49.9	○ 30-45.9		○ 20-29.9		○ ≤ 20
WHITE BLOOD COUNT (total/mm3) (in 1,000s)	○ ≥ 40		○ 20-39.9	○ 15-19.9	○ 3-14.9		○ 1-2.9		○ ≤ 1
GLASGOW COMA SCORE (GCS): Score = 15 minus actual GCS									
A Total ACUTE PHYSIOLOGY SCORE (APS): Sum of the 12 individual variable points									
Serum HCO ₃ (venous-mMol/L) [Not preferred, use if no ABGs]	○ ≥ 52	○ 41-51.9		○ 32-40.9	○ 22-31.9		○ 18-21.9	○ 15-17.9	○ ≤ 15

B AGE POINTS:

Assign points to age as follows:

AGE(yrs)	Points
≤ 44	0
45-54	2
55-64	3
65-74	5
≥ 75	6

C CHRONIC HEALTH POINTS

If the patient has a history of severe organ system insufficiency or is immuno-compromised assign points as follows:

- a. for nonoperative or emergency postoperative patients — 5 points
or
b. for elective postoperative patients — 2 points

DEFINITIONS

Organ Insufficiency or immuno-compromised state must have been evident prior to this hospital admission and conform to the following criteria:

LIVER: Biopsy proven cirrhosis and documented portal hypertension; episodes of past upper GI bleeding attributed to portal hypertension; or prior episodes of hepatic failure/encephalopathy/coma.

CARDIOVASCULAR: New York Heart Association Class IV.

RESPIRATORY: Chronic restrictive, obstructive, or vascular disease resulting in severe exercise restriction, i.e., unable to climb stairs or perform household duties; or documented chronic hypoxia, hypercapnia, secondary polycythemia, severe pulmonary hypertension (>40mmHg), or respirator dependency.

RENAL: Receiving chronic dialysis.

IMMUNO-COMPROMISED: The patient has received therapy that suppresses resistance to infection, e.g., immuno-suppression, chemotherapy, radiation, long term or recent high dose steroids, or has a disease that is sufficiently advanced to suppress resistance to infection, e.g., leukemia, lymphoma, AIDS.

APACHE II SCORESum of **A** + **B** + **C****A** APS points**B** Age points**C** Chronic Health points

Total APACHE II

be desirable to use the acute physiologic portion, the APS score based on 12 physiologic variables, of APACHE II alone. The impact of age and chronic health impairment should be evaluated separately. Elsewhere it has been demonstrated that APACHE II is somewhat more precisely related to hospital survival than the original APACHE (Knaus *et al.*, submitted for publication, 1984a).

APACHE II and resource costs

This section examines the relationship between APACHE II on ICU admission and subsequent total resource costs of treatment over the entire course of the ICU stay. It demonstrates that the APS is highly significantly related to variations in individual cost of care across all patients and that this relationship is robust to the inclusion or exclusion of a large number of diagnostic variables. The importance of severity in explaining costs within specific diagnoses is then evaluated. For one of these diagnoses, interhospital variations in observed costs and efficiency are examined. Finally, whether these interhospital differences are averaged out when all diagnoses are included is analyzed by comparing cost as observed, as adjusted for diagnosis, and as adjusted for diagnosis and severity.

The data base is a newly completed multihospital data base. Information was collected daily in the intensive care units on 200 to 500 consecutive admissions to 12 hospitals, and on almost 2,000 consecutive admissions in a 13th (GW) hospital. Most of these hospitals are medical center teaching hospitals with an average size of about 500 beds. All of the data were collected during 1982, except for one hospital in which the data were collected from 1979 to 1981. In aggregate, there are 5,815 ICU admissions in the data base, of whom 25 are missing some of the resource cost data and are excluded from this analysis.

The cost measure used in this analysis is the Therapeutic Intervention Scoring System (TISS) which was originally developed at Massachusetts General Hospital (Cullen *et al.*, 1974). TISS is an activity analysis measure which groups patient care activities into approximately 75 different items and then assigns relative weights, ranging from 1 to 4, to the items according to ICU nursing time and effort. TISS is intended to directly measure the ICU labor effort, although previous research has indicated that it is highly correlated (0.6 to 0.8) with charges for ancillary services consumed by ICU patients during their ICU stay. (Wagner, Wineland, and Knaus, 1983).

TISS is measured by determining which of the 75 items a patient received during a specific time-period, usually 24 hours, and adding up the corresponding weights. Conventional wisdom holds that a full-time experienced ICU nurse can produce approximately 40 TISS points. A postcoronary artery bypass graft patient requires aggressive management during the

first 24 hours after surgery and typically receives 35 TISS points during the first ICU day. In contrast, patients admitted to medical center ICU's solely for monitoring typically receive a minimum of 10 to 15 TISS points per day because of standard operating procedures.

Based on data from one hospital, the estimate of the resource cost of producing a TISS point was \$60 in 1979 prices.

In this analysis, TISS was measured each day on every patient, and each patient's total TISS over the entire length of the ICU stay was summed. The aggregate mean was 90.0 with a standard deviation of 156.8. The aggregate distribution of admissions across resource costs is illustrated in Figure 2. As in any analysis of individual patient costs, a few extraordinarily expensive patients can substantially influence aggregate means. In order to limit the impact of any individual patient on the subsequent analyses, all individual observations were truncated at 350 TISS points which would normally correspond to 10 days of intense ICU care. These extraordinarily expensive patients, who account for 4 percent of all admissions and 16 percent of total costs, remain included in the analysis. Truncating all of the high-cost patients to 350 TISS points reduced the mean to 75.6 and the standard deviation to 85.6.

Figure 3 illustrates the distribution of these patients across the range of severity of illness at ICU admission as measured by the APS of APACHE II at ICU admission. Figure 4 reports the mean TISS points during the ICU stay for the same severity of illness ranges. Average cost increases strongly up to the 20 to 24 APS point group, after which average cost declines moderately. The mild reduction in costs above 24 APS points occurs partly because of increased death rates at the higher severity levels and partly because of increased impact of the truncation of costs for outliers.

The multivariate analyses presented below demonstrate that the strong relationship between the APS of APACHE II and cost illustrated in Figure 4 is robust to the inclusion or exclusion of a number of diagnostic categories.

The causal model underlying the multivariate analysis is the hypothesis that the more severely ill the patient, the more extensive the therapy. Age, failing chronic health status, whether post-operative or not, and the principal ICU admission diagnosis are also expected to be important determinants of subsequent cost of intensive care. Another potential factor is the possibility of interhospital differences in the efficiency of ICU care. The estimated equation is of the form:

$$\text{Cost} = F(\text{APS, age category, surgical status, diagnostic category, hospital})$$

Figure 2
Therapeutic intervention scoring system (TISS) distribution

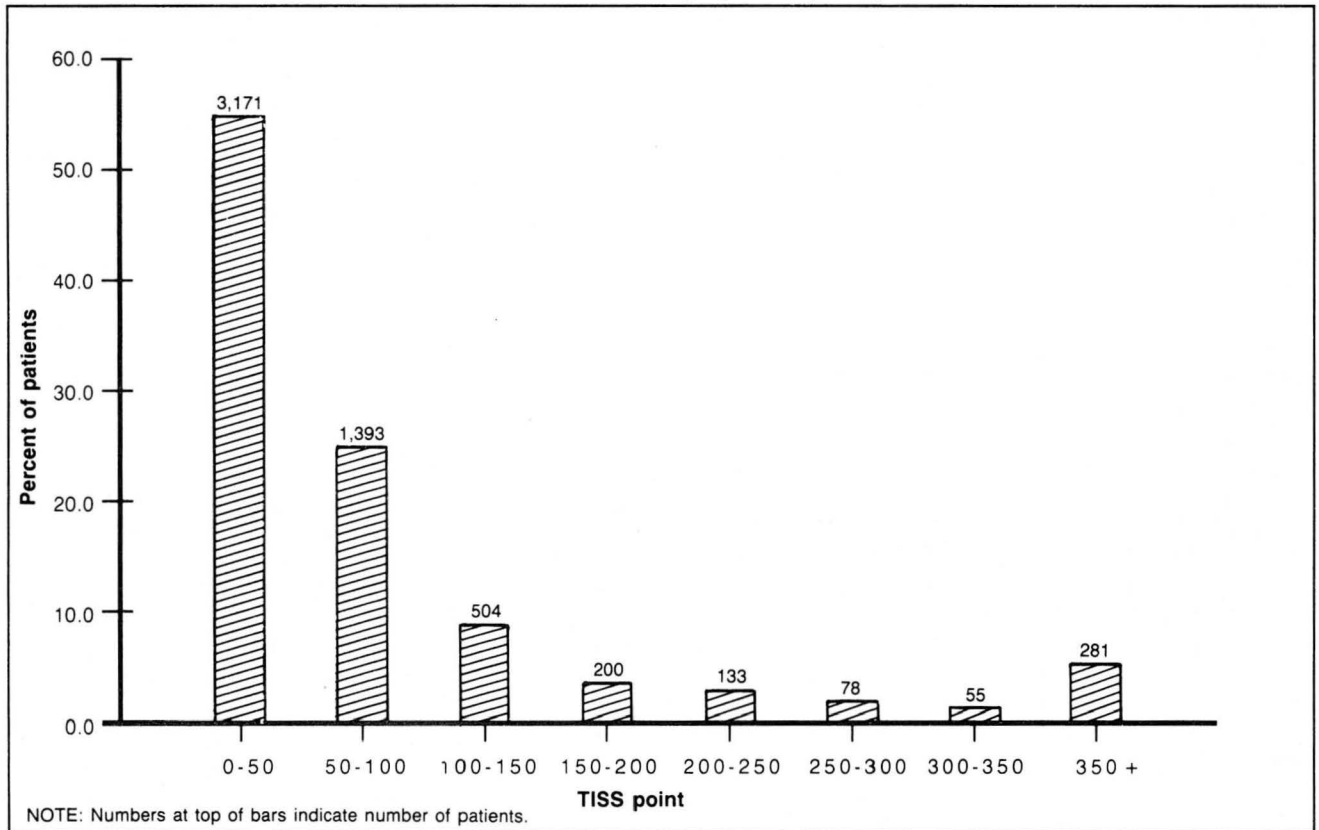


Figure 3
Acute physiology score (APS) distribution

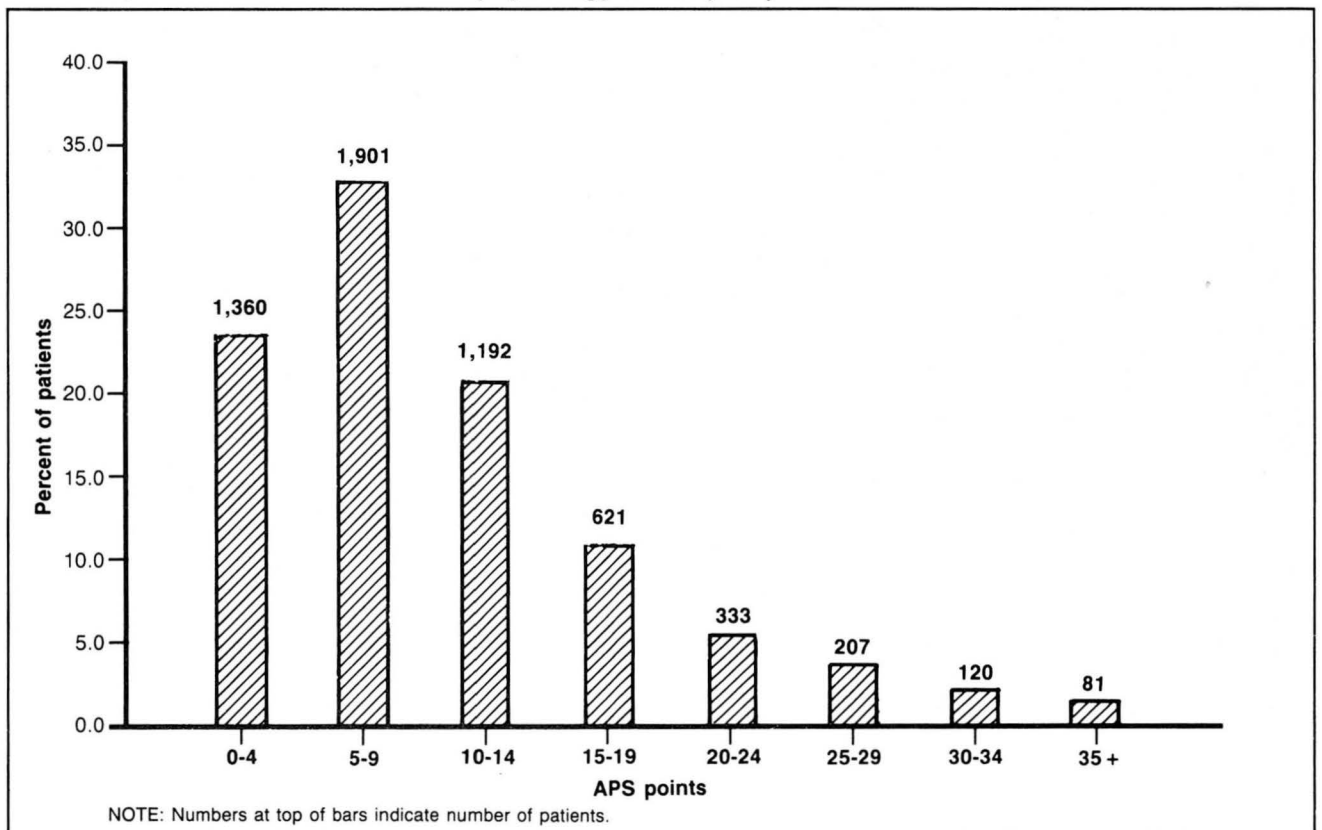


Figure 4
Acute physiology score (APS) and mean therapeutic intervention scoring system (TISS) distribution

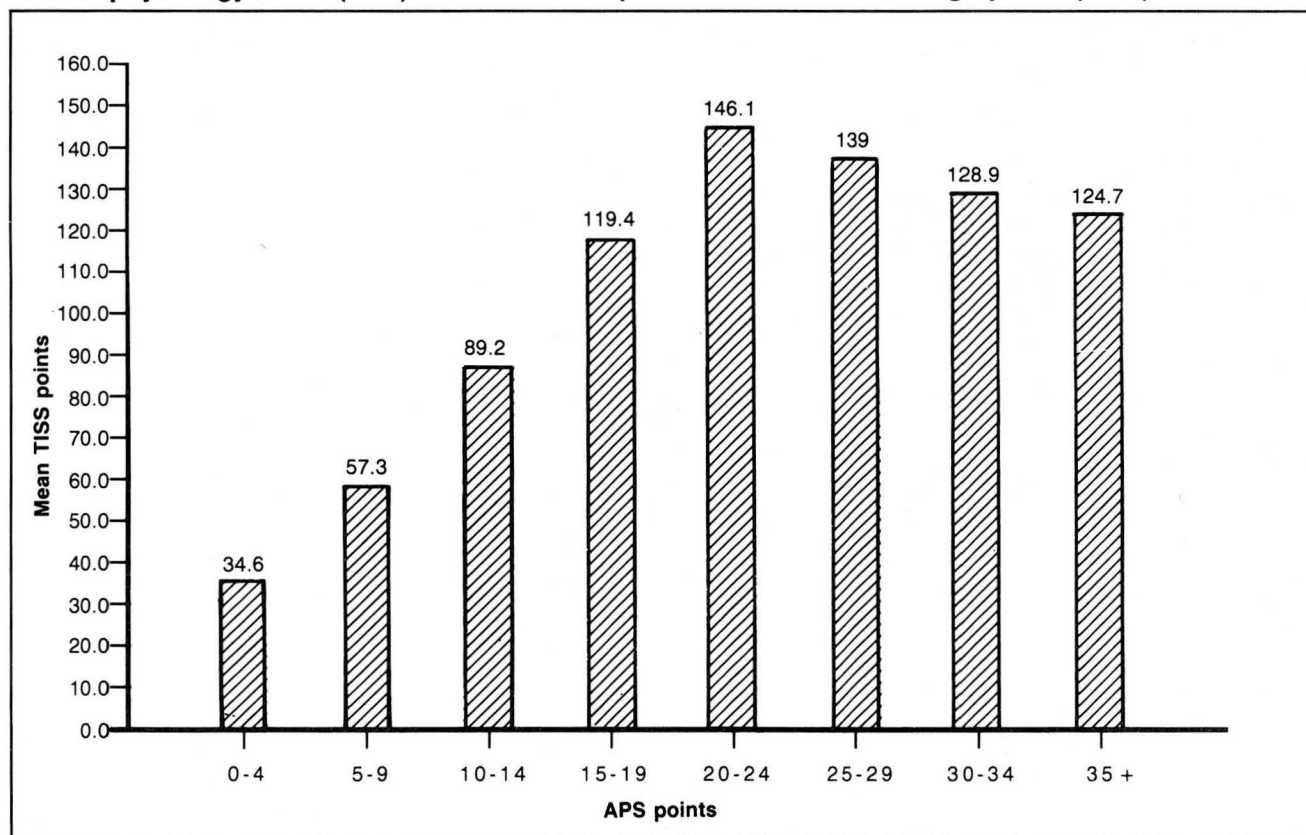


Table 1 reports 3 alternative specifications of the equation estimated with ordinary least squares regression. The first equation includes all variables, Equation 2 deletes the APS score, and Equation 3 includes only APS, age categories, and operative status. All independent variables except the APS score are dichotomous, most in groups of mutually exclusive categories. The reference group for the dichotomous variables is a patient who was admitted with the principal diagnosis of intracranial bleeding, under age 45, post elective surgery, not in chronic failing health, and treated in Hospital 11.

The principal result of the regression equations is the demonstration that interpatient variations in total ICU costs are strongly dependent on acute physiologic derangement shortly after ICU admission. Comparison of the 3 equations reveals that the APS alone accounts uniquely for 38.6 percent of the explained variation, and the 24 diagnostic variables and 12 hospital identifiers together account for only 24 percent of the explained variation. The APS coefficient is very robust to the inclusion or exclusion of the 36 diagnostic variables and hospital identifiers. In contrast, the size of the coefficients on many of the diagnostic variables changes substantially when severity of illness is excluded from the equation.

An alternative specification of the equation, the double logarithmic functional form, resulted in little change in explanation of observed cost and little change in the coefficients.

A number of the other variables merit brief discussion. First, the age pattern is consistent with expectations, with older patients receiving more care after adjustment for severity. The reduced coefficient for those 75 years of age or over is believed to be because of a much smaller portion of elective surgery patients over 75 years of age. Thus, this is probably a result specific to ICU use and not applicable to general hospital utilization. The positive coefficients and strong significance of the nonoperative and post-emergency surgery patients are consistent with clinical expectations for these acutely ill patients, as is the pattern of cost differences across diagnoses. Diagnostic categories that tend to respond relatively quickly to ICU therapy, such as diabetics and drug overdose patients, have substantial negative coefficients and low costs; those that respond poorly or slowly, such as septic shock or some respiratory patients, have positive coefficients and higher than average costs.

One might expect that some of these variable coefficients are biased by the artificially truncated hospital stay of the deceased. Ten percent of these patients died in the ICU, some of them very quickly. When a variable measuring death in the ICU was included, a strongly significant positive coefficient was found, implying the ICU deaths received substantially more care than would otherwise be predicted from admission data. The APS coefficient was not substantially influenced.

Table 1
Regression analysis of APACHE¹ II and resource cost per case

Variable	Equation 1	Equation 2	Equation 3	Variable	Equation 1	Equation 2	Equation 3
Acute physiology score (APS12)	3.81* (25.16)	—	3.98* (28.20)	Other cardiovascular	24.38* (4.31)	14.85 (2.50)	—
Age groups**				Respiratory infection	60.79* (8.39)	59.36* (7.78)	—
45-54 years	8.36 (2.55)	9.92 (2.87)	13.00* (3.97)	Allergy (asthma)	- 6.00 (- 0.44)	- 18.55 (- 1.30)	—
55-64 years	10.28* (3.44)	13.51* (4.30)	15.56* (5.29)	Other respiratory	21.54* (3.94)	9.81 (1.71)	—
65-74 years	18.84* (6.12)	22.81* (7.04)	24.28* (8.08)	GI bleeding	16.89 (2.48)	3.15 (0.44)	—
75 years or over	7.22* (1.97)	12.85* (3.34)	13.29* (3.72)	GI perforation	28.72* (3.41)	19.79 (2.23)	—
Severe chronic health	- 7.28 (- 2.70)	- 2.58 (- 0.91)	- 2.52 (- 0.94)	GI infection	32.72 (2.83)	27.02 (2.22)	—
Operative status**				Other gastrointestinal	15.74 (2.36)	8.20 (1.16)	—
Non-operative	11.44* (3.61)	29.25* (9.00)	7.22* (3.00)	Renal	2.76 (0.35)	4.98 (0.59)	—
Emergency surgery	25.92* (7.18)	38.93* (10.34)	26.51* (8.08)	Metabolic	- 23.64 (- 2.83)	- 22.08 (- 2.51)	—
Admission diagnostic categories**				Hospital identifiers			
Head trauma	9.26 (1.20)	1.98 (0.24)	—	Hospital 1	- 5.03 (- 0.75)	- 23.02* (- 3.29)	—
Drug overdose	- 21.97 (- 2.77)	- 40.83* (- 4.92)	—	Hospital 2	5.27 (0.81)	1.99 (0.29)	—
Craniotomy-neoplasm	- 4.84 (- 0.74)	- 19.12 (- 2.78)	—	Hospital 3	3.77 (0.58)	3.14 (0.46)	—
Other neurologic	- 7.62 (- 1.14)	- 16.22 (- 2.30)	—	Hospital 4	13.85 (1.80)	6.19 (0.76)	—
Post arrest	9.16 (1.34)	30.25* (4.23)	—	Hospital 5	- 5.95 (- 0.76)	- 14.33 (- 1.74)	—
Hemorrhagic shock	26.93* (3.03)	33.51* (3.58)	—	Hospital 6	10.01 (1.56)	7.22 (1.06)	—
Rhythm disturbance	- 9.03 (- 0.99)	- 29.44* (- 3.09)	—	Hospital 7	11.48 (1.55)	6.00 (0.77)	—
Multiple trauma	20.05 (2.88)	11.86 (1.62)	—	Hospital 8	4.65 (0.60)	- 0.83 (- 0.10)	—
Sepsis	33.37* (4.47)	57.23* (7.34)	—	Hospital 9	11.44 (1.49)	10.78 (1.34)	—
Congestive heart failure	16.03 (2.17)	13.53 (1.73)	—	Hospital 10	- 0.70 (- 0.12)	3.82 (0.63)	—
Hypertension	- 23.48 (- 2.27)	- 32.64* (- 3.01)	—	Hospital 12	- 6.40 (- 0.81)	12.09 (1.47)	—
Peripheral vascular	- 0.36 (- 0.06)	- 11.37 (- 1.82)	—	Hospital 13	30.43* (4.74)	34.32* (5.08)	—
Open heart surgery				Intercept	1.83 (0.24)	35.68* (4.58)	15.47* (6.08)
Coronary artery bypass graft (CABG)	23.17* (4.14)	18.46* (3.13)	—	R-squared	.221	.135	.169
Valve repair	52.00* (4.37)	53.15* (4.24)	—	F ratio	37.07*	20.90*	146.5*
				N	5790	5790	5790

*Significantly ($p < .001$) different from 0.0. (t-ratios are in parentheses).

**Reference category for categorical variables is a patient under 45 years of age, admitted to the ICU in Hospital 11 after elective surgery, whose principal admission diagnosis was intracranial bleeding.

¹Acute physiology and chronic health evaluation.

NOTE: The dependent variable is measured in therapeutic intervention scoring system (TISS) points, a nursing intensity scale whose total costs were approximately \$60 per unit in 1979.

The important reimbursement question is—how much does severity of illness add to explanation of interhospital variations in cost? Comparison of the hospital coefficients in equations 1 and 2 indicates that for some hospitals severity of illness is an important determinant of the cost of care. Some hospitals that appear to be less expensive, (hospitals 1, 4, and 5) are substantially less efficient if one takes into account their low severity of illness. Others, particularly hospital 12, would be substantially more efficient than they appear if severity is excluded from the analysis.

It would be desirable to analyze interhospital variations in cost separately for each specific diagnostic category, but this data base does not contain enough observations for reliable interhospital multivariate analysis within most diagnostic categories. It is possible, however, to examine average costs across severity within specific diagnostic categories. Three diagnoses were selected that are relatively frequent and on the lower spectrum of severity of illness. They are more similar to the severity distribution to be expected across a sample of hospital admissions than most diagnoses in this data set.

Table 2 reports the distribution and mean cost within severity range for drug overdose patients, patients admitted to the ICU after peripheral vascular surgery, and diabetics. There is a consistent pattern of substantially increasing cost within each of the three diagnoses that mirrors the aggregate relationship illustrated in Figure 4.

Table 2
Cost by severity within three diagnoses

Diagnoses	APS range ¹	Number	Mean TISS 2 Points	Standard error of mean
Drug overdose				
	0-4	43	9.8	0.91
	5-9	44	12.8	1.10
	10 +	62	64.5	10.64
Peripheral vascular surgery				
	0-4	164	29.7	2.34
	5-9	218	43.5	3.43
	10 +	104	98.8	9.79
Diabetics				
	0-4	11	21.5	3.60
	5-9	23	25.4	4.58
	10 +	84	61.7	9.10

¹Acute physiology score.

²Therapeutic intervention scoring system.

The 468 admissions in the peripheral vascular surgery group are sufficient to examine interhospital differences in cost with some precision. Table 3 reports a multiple regression analysis of this specific diagnostic category in which total ICU cost is assumed to be dependent on severity of illness, age,

and pre-existing severe chronic health. The coefficient on the APS variable is substantial and highly significant. The equation was then used to forecast predicted average costs for the 8 hospitals that have more than 20 patients in this diagnostic category.

Figure 5 plots the observed cost and the predicted cost based on the regression analysis. Within this narrow disease category peripheral vascular surgery, there is large variation across hospitals in observed cost and considerable difference in severity of illness. This leads to substantial differences in predicted costs of ICU care. Dividing predicted costs by observed costs yields efficiency ratios that range from 2.4 to .6, where the average efficiency in this sample as a whole is normalized to 1.0. Thus, within this specific diagnostic category there are wide interhospital differences in the severity of patients at admission and in the efficiency of ICU care.

Are these disease-specific differences averaged out across all diseases? Figure 6 indicates they are reduced but not eliminated. Figure 6 plots the mean cost per case computed three ways for each hospital. The first value for each hospital is the observed average cost for all patients, which ranges from 49 to 102 TISS points. Second, a predicted average cost based on equation 2 in Table 1, is reported for each hospital, though the hospital variables were not used in the projection. This is a predicted cost based on inter-hospital variations in diagnostic mix but not severity.

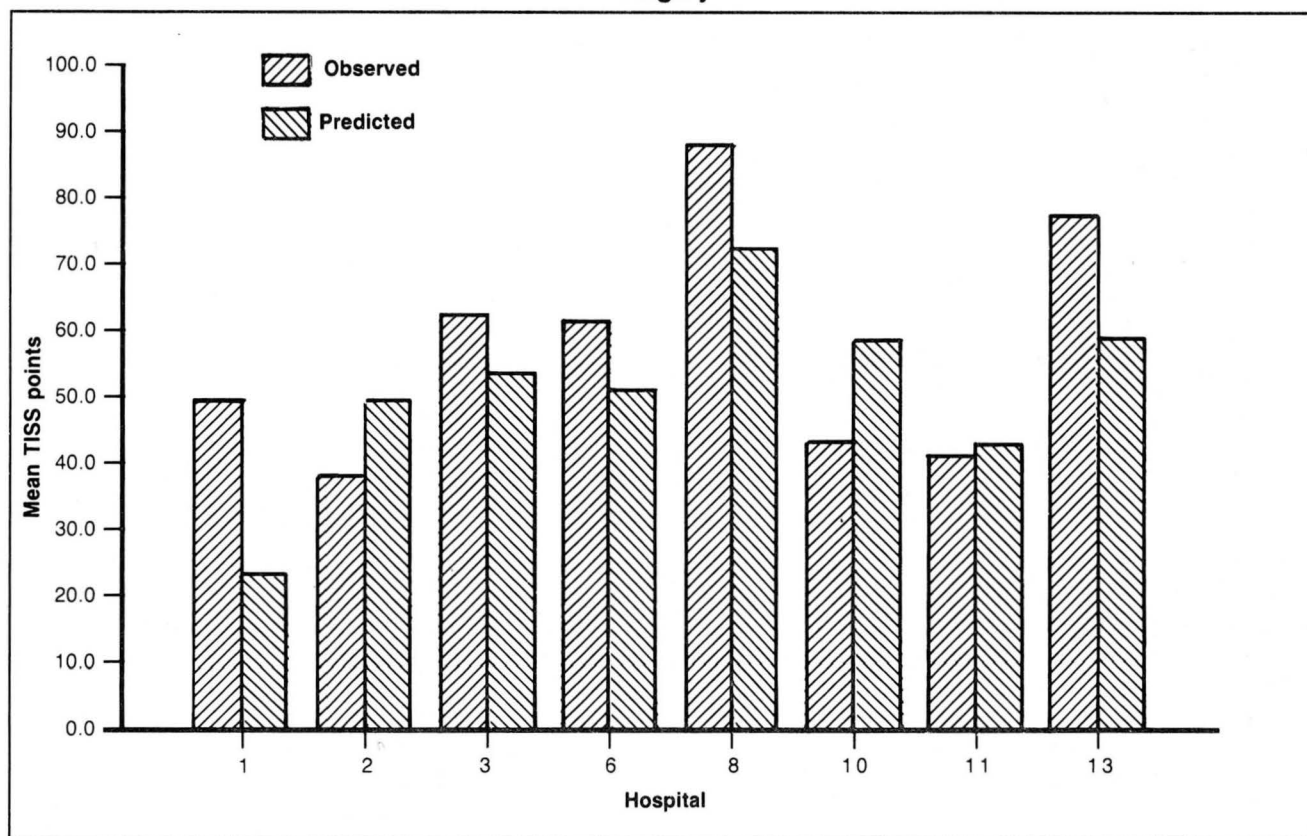
Table 3
Regression analysis of severity of illness and cost among post-surgical peripheral vascular patients

Variables	
Acute physiology score	5.88* (10.64)
Age groups	
45-54	- 1.15 (- 0.07)
55-64	8.44 (0.65)
65-74	17.96 (1.39)
75 years or over	9.43 (0.68)
Severe chronic health	- 26.57 (- 2.56)
Intercept	1.08 (0.08)
R-squared	.217
F ratio	22.13*
N	486.

*Significantly ($p < .001$) different from 0.0. (t-ratios are in parentheses).

NOTE: The dependent variable is measured in therapeutic intervention scoring system (TISS) points, a nursing intensity scale whose total costs were approximately \$60 per unit in 1979.

Figure 5
Observed and predicted therapeutic intervention scoring system (TISS) for peripheral vascular surgery



Third, a predicted cost was computed based on equation 1, severity as well as diagnostic mix, again excluding the hospital coefficients from the computation of predicted costs.

For 3 hospitals there is little difference among the 3 cost measures, but for the other 10 hospitals there are substantial differences between observed costs and severity-predicted costs based on severity with efficiency ratios ranging from .8 to 1.4. Thus, the efficiency ratios are compressed toward 1.0 by including all patients, but differences persist. One hospital is substantially (40 percent) more expensive than expected, controlling for severity of illness, and 2 hospitals are substantially (20 percent) more efficient. The other 7 hospitals average a 10-percent differential between observed and predicted cost.

If we turn to a different question and compare the costs predicted based on severity with costs predicted from diagnoses alone, in aggregate the average magnitude of the differential is similar. For individual hospitals, however, the difference between the two costs are often quite substantial. Moreover, the magnitude of this differential for individual hospitals is markedly changed from the difference between observed costs and severity-predicted costs described above. For example, severity adjusted reimbursement would increase hospital 12's reimbursement by approximately 12 percent, but reimbursement based

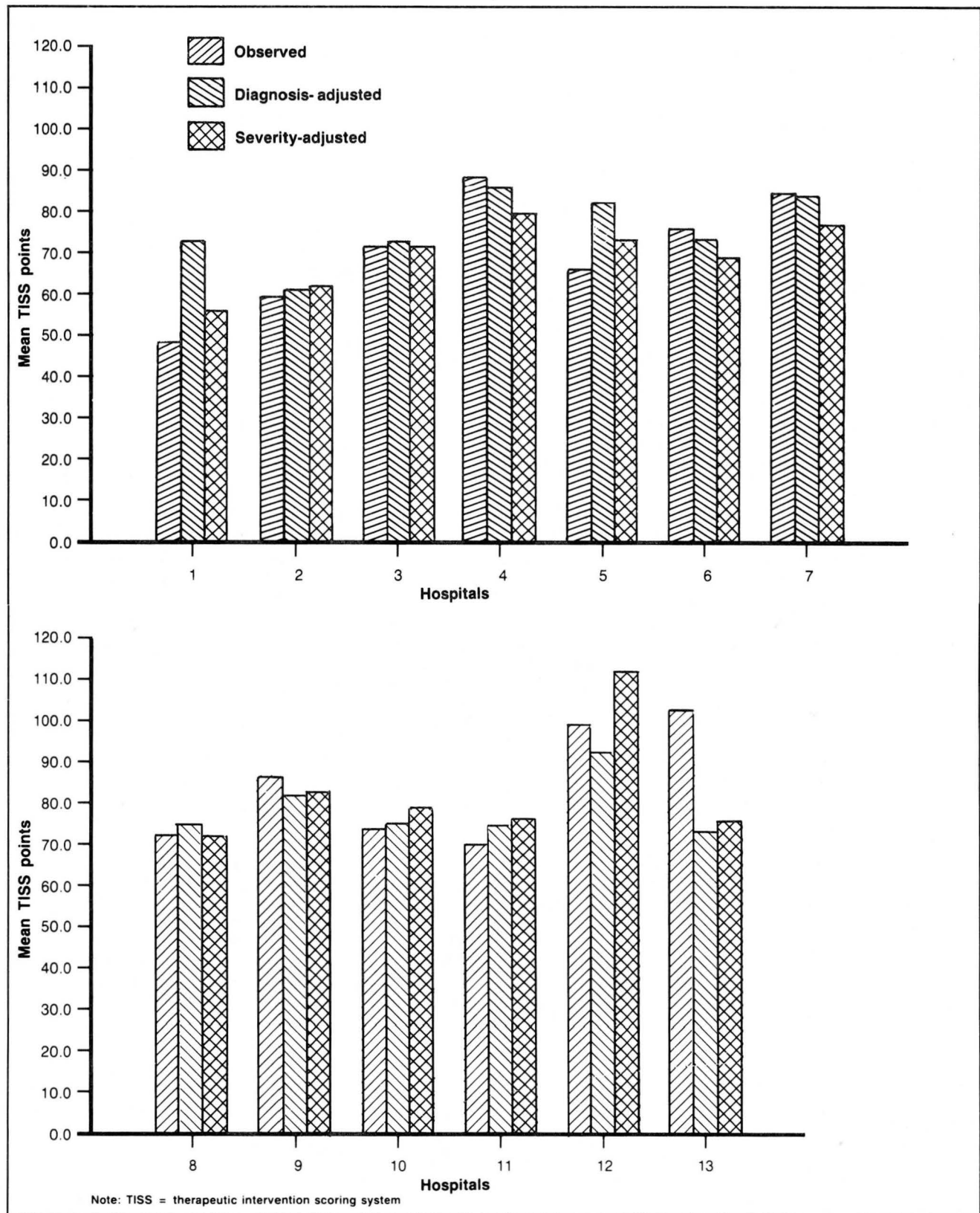
on the equation that excludes severity would reduce its revenue by 12 percent.

Predicted costs based on diagnoses without severity substantially reduce interhospital variation in average cost per case. The standard deviation in observed cost across the 13 hospitals is 14.9. The comparable standard deviation in predicted costs based on diagnoses is only 7.6. In contrast, the standard deviation based on severity of illness and diagnoses is 12.4. The latter number seems more consistent with clinical judgement based on onsite inspection. Several of these hospitals have extraordinarily severe case mixes requiring extensive therapy, and others have larger numbers of low-risk monitor patients who may not need to be in an ICU. This suggests, but does not prove, that adjustment for diagnoses masks important interhospital differences in case mix.

Research needs

It is important for the reader to clearly understand that none of the empirical results presented here are directly applicable to prospective reimbursement questions. The data samples, being only ICU admissions, are biased samples of most DRG's. The diagnostic categories are ICU admission diagnoses rather than hospital discharge diagnoses. These categories identify patient groups that are diagnostically homogeneous

Figure 6
Observed and predicted cost



for ICU physicians. Some of these diagnostic categories are more narrowly defined than the corresponding DRG's and for others the reverse is true. An example of the former is that we divide brain surgery patients into those undergoing the surgery because of cancer versus patients with intracranial hemorrhage and stroke. An example of the latter is that all of the post-coronary artery bypass graft patients are in the same category, regardless of age and prior complicating condition or comorbidity.

The most important difference between this data and hospital discharge DRG data is that the APS measures are taken within 24 hours of ICU admission, before the outcomes occur. In addition, the cost measures, though more accurate than charges or hospital accounting data, cover only the ICU stay. The hospitals sampled were not a representative sample of all hospitals. They were mainly large teaching hospitals and have less interhospital variation than one would expect across all hospitals.

The method illustrated here, however, could be applied to a large sample of hospital admissions. The first task would be to measure the relationship between the APS of APACHE II at hospital admission and resource cost over the entire hospital stay, within DRG categories. It is reasonable to expect the APS measure to capture severity of illness differences for most common diagnoses. The only patients systematically excluded from the data reported here are psychiatric, obstetric, suspected heart attack, burn victims, and children. APS may be sufficiently precise in some of these patient groups for reimbursement purposes, though other medical research has developed disease-specific severity measures that are more closely attuned to the clinical questions (Pozen *et al.*, 1984; Goldman *et al.*, 1982; Fuchs and Scheidt, 1981; Killip, 1972; Mulley *et al.*, 1980; Feller *et al.*, 1980, Yeh *et al.*, 1984).

Research conducted by others has established that retrospective review of medical charts can yield accurate data on longer versions of the APS score (U.S. General Accounting Office, 1983, Multnomah PSRO, 1983). Another research group has measured the longer APS using 34 physiologic variables on samples of non-ICU patients in one hospital and found the APS significantly associated with interpatient variations within DRG categories in cost-adjusted charges (Coulton, 1984). Analysis of our own data indicates that the first physiologic parameters available at ICU admission are as sensitive as the APS defined by the worst physiologic values over a 24-hour period.

The second, and equally important, task would be to determine whether there are substantial differences in costs within DRG's across hospitals, and whether these differences are substantially associated with severity differences.

It seems likely that a number of hospitals already have the data items necessary for the APS in computerized records which can be linked to the discharge abstract and hospital bill. Several of the data items are included but rarely used in the Professional Activities Study (PAS) discharge data set. Most of the

others are measured by computer-based instruments which produce a small computer printout as the laboratory report. Many hospitals probably store this data in electronic media. In the long run there would be a very low marginal cost per case in using a machine-readable output from these laboratory tests.

The distribution of hospital admissions across the APS is likely to be substantially different than for ICU patients. One would expect a large majority of patients in most DRG categories in most hospitals to be in the 0-4 or 5-9 (low severity) point range at hospital admission. The measure, however, is sensitive enough even in this range to pick up the physiologic consequences of many comorbid conditions and secondary diagnoses. The central question for reimbursement purposes will be, how large is the cost differential for patients in the 5-9 point range or over 10 APS point range, and how unequally are these more severely ill patients distributed across hospitals? The results on peripheral vascular surgery patients, drug overdose, and diabetics suggest that the APS could be strongly related to interhospital variations in costs within DRG's at the lower severity ranges.

The APS of APACHE II could easily be integrated with DRG's in two different ways. First, it could simply be viewed as a multiplier for each individual cell. The second approach would be to ascend the DRG decision tree one level, to the point at which most DRG's are divided on the basis of the presence of comorbidity, complications, secondary diagnoses, or age over 70. This split, instead, could be systematically based on APS level, modified or in combination with age. Evaluation of these and other options would require detailed analysis.

Potential advantages

There are a number of potential advantages possible from gathering objective physiologic data close to hospital admission on virtually all hospital patients. Many of the advantages would lie in the enhanced ability to evaluate the interhospital homogeneity of DRG's and to evaluate the impact of DRG reimbursement, whether or not the DRG system is ultimately adjusted for severity of illness. An independent measure of severity of illness would provide useful information for assessing hospitals' complaints about the validity of the DRG system. The APS of APACHE II would also provide the capacity to monitor the extent to which hospitals begin to triage sicker patients to local public hospitals or other tertiary care hospitals. Public hospitals are concerned about increased adverse selection of their patient mix as a consequence of prospective reimbursement.

Another important application of the information embodied in the APS would be to improve the precision of the analysis of the efficiency of individual physician's hospital practice. It appears that the principal mode by which prospective reimbursement will lead to real resource savings is in setting standards for cost of care for various DRG categories. When hospital medical directors or department chairmen confront their high-cost physicians, much of the

discussion is going to be an argument on whether that particular physician's patients within that DRG are more severely ill than the average. Objective information about severity of illness which is not sensitive to individual physician judgment or aggressiveness of care could substantially improve those discussions.

From the viewpoint of classic evaluation methodology as well as clinical acceptability, it would be desirable to base the diagnostic categorization of patients on information available at hospital admission and not after hospital discharge (Campbell and Stanley, 1963). A substantial portion of the uncertainty and difficulty of medical decision making is assumed away by using retrospective discharge diagnoses. Hindsight is usually more accurate but less useful than foresight, and it provides less incentive for efficiency and quality of care. This distinction would be less important if diagnoses were as objective, distinct, and homogeneous as many nonphysicians believe. Many diagnoses include a heterogeneous group of conditions that overlap with other diagnoses, perhaps depending as much on individual variation across physicians as across patients, even in the absence of financial incentives.

Thus it would be quite feasible to combine information on severity of illness at admission with admission diagnostic information in a new patient categorization system (Young *et al.*, 1982). APACHE II would also provide an enhanced ability to evaluate interhospital variations in the quality of care, and the regionalizing impact of prospective reimbursement (Knaus *et al.*, submitted for publication, 1984b). Regardless of reimbursement applications, it appears that the APACHE II system of prognostic stratification will begin to play an important role in research in the clinical and biomedical medical sciences. (Amos *et al.*, 1982; Feinstein, 1983; Meakins *et al.*, 1984; Kurek *et al.*, 1984)

Summary and conclusions

There is substantial evidence for interhospital differences in severity of patients after controlling for DRG's. Measurement of objective physiologic parameters at hospital admission would be an accurate and appropriate method of assessing the magnitude of these differences and planning a policy response. One method that could be used is APACHE II, a classification system for severity of illness that has just undergone national validation.

The Acute Physiology Score (APS) of APACHE II is based on 12 objective physiologic measurements, most of which are routinely measured on a large majority of hospital patients shortly after admission. This article has demonstrated that the APS measured shortly after admission to an intensive care unit has a strong and stable relationship with resource cost of subsequent intensive care. These results are demonstrated across all diagnoses and within diagnoses for a national sample of 5,790 intensive care unit admissions at 13 large hospitals.

The strength of APACHE II is its timing and objectivity of measurement, clinical acceptance,

breadth of diagnostic coverage, and robust statistical performance in predicting costs and survival for ICU patients. Data collection and analysis designed for reimbursement purposes on large samples of hospital patients has not yet begun. If appropriate computerized data bases can be located, the analysis could be completed in 6 months.

There are a number of other important policy issues regarding the impact of Medicare hospital reimbursement which require serious evaluation. Perhaps the most important to the Federal budget is the question about appropriate DRG prices (Lave, 1984). The average cost pricing implicit in past hospital accounting conceals important interpatient subsidies (Williams *et al.*, 1982; Wagner, Wineland, and Knaus, 1983). If surgery is overpriced relative to medicine, will we have even more surgery? What is the impact of prospective reimbursement on the beneficiaries? Because DRG reimbursement is likely to influence discharge diagnoses, accurate answers to these and other questions will require original data collection in a number of settings.

Implementation of the Tax Equity and Fiscal Responsibility Act of 1982 (TEFRA) and prospective reimbursement based on DRG's has caused profound changes in hospitals across the country. For the first time in many years, hospital managers must be concerned about the cost effectiveness of the medical practice in their hospitals. In response, they are intensely reviewing the use of resources in their own institutions, attempting to correct the most blatant overuse, and developing systems for predicting DRG categories at hospital admission for prospective cost control. This seems likely to result in stronger control of hospital medical practice by hospital medical directors and department chairmen. Though it is still far too early to assess the impact of DRG's, the changes are likely to have positive impacts on costs and quite possibly quality of care.

It is suspected that a large number of the 468 DRG categories, particularly many of the elective surgery categories, cannot be substantially improved upon without initiating careful evaluations of the indications for surgery. There has been extensive research documenting large variations in surgical rates across areas, with the implication that too much surgery is done (Wennberg and Gittelsohn, 1973; Wennberg *et al.*, 1984; Roos and Roos, 1981). This seems to be an important area in need of substantially more work, for the benefit of the patients as well as the taxpayer and those paying for private health insurance.

There is a curious contrast between research on medical care and research in other areas of economic activity. Usually 50 to 80 percent of research is done to develop new processes, primarily to produce existing products at lower cost (Mansfield, 1980). In contrast, in the highly research-intensive medical area, almost all of the research is orientated toward producing new products, or new cures. Little of the research is designed to lower the cost of producing the same products. Prospective reimbursement will substantially increase the demand for such information. Because of

ethical constraints on research on human beings, large sample-size requirements, and a limited number of experienced researchers, the supply of such information will lag far behind.

In order to save 10 to 30 percent of the annual \$50 billion Medicare expenditure on hospitals and the \$122 billion aggregate expenditure on hospitals (without adversely affecting patients), it will be necessary to substantially expand cost effectiveness and medical care evaluation research and publication. Others have previously suggested a research effort of 2 tenths of 1 percent of medical care expenditures would be appropriate (Relman, 1980; Bunker and Fowles, 1982). For the \$322 billion dollar health care industry, that amounts to \$644 million for evaluation, with \$244 million of the total focused on hospital care.

Acknowledgments

The author is indebted to Dr. William Knaus, Mr. Julian Pettengill, and an anonymous referee for constructive suggestions on a previous version of this paper.

References

- Amos, R. J., Amess, J. A. L., Hinds, C. J., and Mollin, D. L.: Incidence and pathogenesis of acute megaloblastic bone-marrow change in patients receiving intensive care. *Lancet*, II:835-838, 1982.
- Bunker, J. P., and Fowles, J.: Model for an institute for health care evaluation. Appendix F in *Strategies for Medical Technology Assessment*. U.S. Congress, Office of Technology Assessment, OTA-8-181, Washington. U.S. Government Printing Office, Sept. 1982.
- Campbell, D. T., Stanley, J. C.: *Experimental and Quasi-Experimental Designs for Research*. Rand McNally & Company, 1963.
- Coulton, C., McClish, D., Doremus, H., et al.: Implications of DRG's for intensive care. *Crit Care Med* 12(3):332 (Abstract), 1984.
- Cullen, D. J., Civetta, J. M., Briggs, B. A., and Ferrara, L. C.: Therapeutic intervention scoring system: A method for quantitative comparison of patient care. *Crit Care Med* 2:57-60, 1974.
- Draper, E. A., Wagner, D. P., and Knaus, W. A.: The use of intensive care: A comparison of a university and community hospital. *Health Care Financing Review*. Vol. 3, No. 2. HCFA Pub. No. 03139. Office of Research, Demonstrations, and Statistics, Health Care Financing Administration. Washington. U.S. Government Printing Office, Dec. 1981.
- Feinstein, A. R.: An additional basic science for clinical medicine I: The constraining fundamental paradigms. *Ann Intern Med* 99:393-397, 1983.
- Feller, I., Tholen, D., and Cornell, R. G.: Improvements in burn care, 1965 to 1979. *JAMA* 244:2074-2078, 1980.
- Finkler, S. A.: The distinction between cost and charges. *Ann Intern Med* 96:102-109, 1982.
- Fuchs, R., and Scheidt, S.: Improved criteria for admission to cardiac care units. *JAMA* 246:2037-2041, 1981.
- Garber, A. M., Fuchs, V. R., and Silverman, J. F.: Case mix, costs, and outcomes: Differences between faculty and community services in a university hospital. *N Engl J Med* 310:1231-1237, 1984.
- Goldman, L., Weinberg, M., Weisberg, M., et al.: A computer-derived protocol to aid in the diagnosis of emergency room patients with acute chest pain. *N Engl J Med* 307:588-596, 1982.
- Horn, S. D.: Measuring severity of illness: Comparisons across institutions. *Am J Public Health* 73:25-31, 1983.
- Killip, T.: Problems in myocardial infarction. In Ruseck, H. I., Zohman, B. L. (eds): *Coronary Heart Disease*. Philadelphia. J. B. Lippincott Co., 1972.
- Knaus, W. A., Zimmerman, J. E., Wagner, D. P., et al.: APACHE—Acute physiology and chronic health evaluation: A physiologically based classification system. *Crit Care Med* 9:591-597, 1981.
- Knaus, W. A., Draper, E. A., Wagner, D. P., et al.: Evaluating outcome from intensive care: A preliminary multihospital comparison. *Crit Care Med* 10:491-496, 1982.
- Knaus, W. A., Le Gall, J. R., Wagner, D. P., et al.: A comparison of intensive care in the U.S.A. and France. *Lancet* II:642-646, 1982.
- Knaus, W. A., Draper, E. A., and Wagner, D. P.: The use of intensive care: New research initiatives and their implications for national health policy. *Milbank Memorial Fund Quarterly*, 61:561-583, 1983.
- Knaus, W. A., Draper, E. A., Wagner, D. P., and Zimmerman, J. E.: APACHE II: A Severity of Disease Classification System for Acutely Ill Patients. Submitted for publication, 1984a.
- Knaus, W. A., Draper, E. A., Wagner, D. P., and Zimmerman, J. E.: An Evaluation of Outcome from Intensive Care. Submitted for publication, 1984b.
- Kurek, T., Zaloga, G. P., Chernow, B., et al.: Total serum T4 concentration correlates with severity of illness (APACHE score) in critically ill euthyroid patients. *Clinical Research* 32:251A (abstract), 1984.
- Lave, J. A.: Hospital reimbursement under Medicare. *Milbank Memorial Fund Quarterly* 62:251-268, 1984.
- Mansfield, E.: Basic research and productivity increase in manufacturing. *The American Economic Review* 70:863-873, 1980.
- Meakins, J. L., Solomkin, J. S., Allo, M. D., et al.: A proposed classification of intraabdominal infections: Stratification of etiology and risk, for future therapeutic trials. *Archives of Surgery*, 1984. To be published.
- Mulley, A. G., Thibault, G. E., Hughes, R. A., et al.: The course of patients with suspected myocardial infarction: The identification of low-risk patients for early transfer from intensive care. *N Engl J Med* 302:943-948, 1980.
- Multnomah Foundation for Medical Care. Intensive Care Unit Areawide Study. Portland, OR: Multnomah Foundation for Medical Care, 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 and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Dec. 1982.

Pozen, M. W., D'Agostino, R. B., Selker, H. P., *et al.*: A predictive instrument to improve coronary care unit admission practices in acute ischemic heart disease: A prospective multicenter clinical trial. *N Engl J Med* 310:1273-1278, 1984.

Relman, A. S.: Assessment of Medical Practices. (Editorial) *N Engl J Med* 303:153-154, 1980.

Roos, N. P., and Roos, L. L.: High and low surgical rates: Risk factors for area residents. *Am J Public Health* 71:591-600, 1981.

U.S. General Accounting Office. Human Resources Division: Survey of the Use of Over Utilization of ICUs by Medicare Beneficiaries. Pilot Study, 1983.

Wagner, D. P., Knaus, W. A., and Draper, E. A.: Statistical validation of a severity of illness measure. *Am J Public Health* 73:878-884, 1983.

Wagner, D. P., Knaus, W. A., Draper, E. A., and Zimmerman, J. E.: Identification of low-risk monitor patients within a medical-surgical intensive care unit. *Med Care* 21:425-434, 1983.

Wagner, D. P., Wineland, T. D., and Knaus, W. A.: The hidden costs of treating severely ill patients: Charges and resource consumption in an intensive care unit. *Health Care Financing Review*. Vol. 5, No. 1. HCFA Pub. No. 03154. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Sept. 1983.

Wennberg, J. E., and Gittelsohn, A.: Small area variations in health care delivery. *Science* 18:1102-1108, 1973.

Wennberg, J. E., McPherson K., and Caper P.: Will payment based on diagnosis-related groups control hospital costs? *N Engl J Med* 331:295-300, 1984.

Williams, S. V., Finkler, S. A., Murphy, C. M., and Eisenberg, J. M.: Improved cost allocation in case-mix accounting. *Med Care*, 20:450-459, 1982.

Yeh, T. S., Pollack, M. M., Ruttimann, U. E., *et al.*: Validation of a physiologic stability index for use in critically ill infants and children, *Pediatr Res*, May 1984.

Young, W. W., Swinkola, R. B., and Zorn D. M.: The measurement of hospital case mix. *Med Care* 20:501-512, 1982.