Relationship of Physician Ratings of Severity of Illness and Difficulty of Clinical Management to Length of Stay

Catherine Kelleher

In a concurrent prospective study, medical and surgical residents rated the severity of illness and difficulty of clinical management of each of their patients within 24 hours of admission, and on a daily basis throughout the patient's stay. Data were collected on consecutive admissions resulting in 661 admissions with complete data for analysis. Results indicate that difficulty and severity are correlated, each explaining variations in length of stay (LOS), and together explaining up to 44 percent. Four alternative measures are tested, first-day values, average values over the stay, peak or highest values, and a zero-one measure indicating whether or not the severity or difficulty fluctuated over the stay. First-day and average measures of severity and difficulty explain little variation in LOS; peak and fluctuating measures are highly explanatory. After adjusting for diagnosis-related groups (DRGs), fluctuating severity adds 34 percent, and adjusting for both DRGs and severity, fluctuating difficulty adds 10 percent for a total of 53 percent variance explained. In comparable results, peak severity adds 21 percent, and peak diffi-

---

This study was supported by a grant from the Health Services Research Council of the Bureau of Medical Services, the U.S. Public Health Service Hospital, and intramural funds from the Office of Health Services Research, the Baltimore U.S. Public Health Service Hospital, and the Wyman Park Health System, Baltimore, MD.

Address correspondence and requests for reprints to Catherine Kelleher, Sc.D., M.P.H., M.S., R.N., Johns Hopkins University School of Nursing, 375 Houck, 600 N. Wolfe Street, Baltimore, MD 21287-1316. Dr. Kelleher is Assistant Professor and Director of the M.S.N./M.P.H. joint degree program for the Johns Hopkins University School of Nursing, with a Joint Appointment in Oncology, the Johns Hopkins University School of Medicine. At the time of this research, Dr. Kelleher was Research Associate in the Office of Health Services Research, the Baltimore U.S. Public Health Service Hospital, and the Wyman Park Health System, Baltimore, MD. This article, submitted to *Health Services Research* on April 30, 1991, was revised and accepted for publication on June 24, 1992.

culty 4 percent, for a total of 34 percent variance explained. Findings indicate that difficulty had independent value as a predictor, and the high explanatory power of the fluctuating measures suggests that a third dimension, instability, may be as important as severity and difficulty in explaining LOS.

It is now well recognized that variations in length of acute care hospitalization are only partially explained by differences in case mix when measured by diagnosis-related groups (DRGs) (Cretin and Worthman 1986; Jencks and Dobson 1987; McMahon and Billi 1988; Mackenzie and Willan 1991). It is suggested that this is due to unexplained differences in severity of patient condition and difficulty of clinical management. The use of these latter measures in conjunction with DRGs would be expected to improve the explanatory power of DRGs by adjusting for patients who are sicker and more complex to manage (Thomas, Ashcraft, and Zimmerman 1986; Aronow 1988; Rosko 1988; Iezzoni, Moskowitz, and Daley 1989; Iezzoni, Schwartz and Restuccia 1991; Burns and Wholey 1991).

Studies investigating use of severity measures to explain variations in length of stay and showing that severity adds to the explanatory power of DRGs have included those that have used the Computerized Severity Index (Horn, Sharkey, Buckle, et al. 1991), MedisGroups (Iezzoni, Ash, and Moskowitz 1987), and Patient Management Categories and Staging (Calore and Iezzoni 1987). But, to date, little information is available regarding the relative contributions of severity and difficulty of clinical management as independent factors affecting length of stay. It might be expected that increasing illness severity contributes to the increasing difficulty of clinical management, particularly if severity reflects failure to respond to initial treatment. Yet there are likely to be exceptions: for example, terminally ill patients for which no effective treatment exists may not necessarily be difficult to manage despite the severity of their condition.

This study brings together the concepts of severity and difficulty of clinical management to examine the explanatory power of each, their interrelationship, and the ways in which they work together in explaining variations in length of acute inpatient stay. It is expected that these measures will be significantly correlated, yet make significant independent contributions to the explanation of variations in hospital stay.

METHODS

DATA COLLECTION

The work described here was a concurrent prospective study of acute care inpatients. Medical and surgical residents completed daily checklist ratings of severity and difficulty on 901 consecutive patients admitted to the Baltimore U.S. Public Health Service Hospital between April 20, 1981 and October 2, 1981. Admissions for alcohol detoxification, one-day chemotherapy, and research protocols were excluded as their length of stay patterns were atypical of medical and surgical admissions. Resident ratings were supplemented by data from the patients' hospital discharge abstracts. Data collection was completed one month prior to termination of federal financing when the hospital transferred to the private sector and became the Wyman Park Health System.

The hospital had 135 beds, and offered services similar to those of an acute care general hospital with the exception of pediatrics and maternity care. It had about 4,000 inpatient discharges per year, almost 60 percent from Baltimore City and Baltimore County. Approximately 40 percent of discharges were retirees or dependents of uniformed services personnel; 20 percent were American naval personnel; and the rest were community residents, including individuals on Medicare and Medicaid, and those without insurance.

The medical staff was organized as a closed panel, primarily of U.S. Public Health Service commissioned officers. It was a teaching facility for the Johns Hopkins University and University of Maryland medical schools, and the panel was supplemented by faculty who served as attending physicians. The ratio of residents to attending physicians was 4:1 for medicine, and 3:1 for surgery. Residents were selected to do the daily severity and difficulty ratings because of their first-hand knowledge of patient status, and because the attending physicians were not consistently available on a daily basis. Medical residents worked in senior-junior pairs defined by year of residency training. The senior of the pair did the ratings, with the junior filling in when the senior was off-duty. Surgical residents worked within teams of three or four members; if the resident managing the patient was off-duty, the back-up resident did the ratings.

Of the 21 medical residents who participated, 9 were senior, 4 in their third year, the rest in their second; of the 12 who were junior pair members, 10 were in their first year, and 2 in their second. Of the ten

surgical residents, there were two each in the third–fifth year, and the rest were in their second.

VARIABLES

Residents were asked: "In the past 24 hours, how ill was this patient (not, minimally, moderately, very)?" Ordinal values were assigned to the nominal ratings after data collection: not = 0, minimally = 1, moderately = 2, and very = 3. Four severity measures were constructed, including a first-day score, the rating for the first 24 hours of stay; an all-days score, the average of all daily ratings; peak, the highest rating for the stay; and fluctuating, a binary variable coded "yes" if the illness severity question was rated in an increasing-decreasing sequence or decreasing-increasing sequence. No adjustment was made for number and magnitude of rating reversals, that is, if the maximum or minimum severity was outside the range of the admission and discharge severity measures, the case was coded as having a fluctuating severity.

Difficulty of clinical management was a composite of ratings for complexity, judgment, uncertainty/unpredictability, and standardization/routineness of clinical management. Selection of these four dimensions was based on literature review and study physician consensus (Feinstein 1967; Engelhardt, Spicker and Towers 1979; Fox 1980; Bursztajn et al. 1981; Schoonhoven et al. 1980). Residents rated the following questions:

- In the past 24 hours, how complex was the case, i.e., had many interrelated components (very complex, moderately complex, minimally complex, not complex)?
- In the past 24 hours, how much clinical judgment was required (a great deal, moderate, minimal, none)?
- In the past 24 hours, how much uncertainty/unpredictability was there in the case management (a great deal, moderate, minimal, none)?
- In the past 24 hours, how standardized/routine was the management of the case (very nonroutine, moderately nonroutine, minimally nonroutine, routine)?

As with severity, ordinal values were assigned to the nominal ratings after data collection: none = 0, minimal = 1, moderate = 2, and great deal = 3. The difficulty scale measure was the sum of the four component questions scored as indicated, then divided by 4, producing a range of 0–3. Four difficulty measures were constructed,

including a first-day scale score, an average of the ratings for the first day of stay; an all-days scale score, an average of item averages for all daily ratings; peak, the highest rating for the stay on any one of the items; and fluctuating, a binary variable for which the admission was coded "yes" if at least one of the four items was rated in an increasing-decreasing or decreasing-increasing sequence; no adjustment was made for number and magnitude of rating reversals.

Internal validity of the difficulty scales was tested via item analysis (Cronbach 1951), and internal reliability via Cronbach's alpha (Bohrnstedt 1969). For the first-day, average, and last-day difficulty scales, the corrected item-to-total correlations of the complexity, judgment, uncertainty/unpredictability, and standardization/routineness items were high and similar, with the ranges of .84–.98 providing evidence of internal validity. Alphas for the first-day, average, and last-day difficulty scales ranged from .83–.98. As alpha of .60 is evidence of internal reliability, the scales were internally reliable.

In addition to severity and difficulty, other variables were obtained from the residents. These included admission status, admission type, hospitalized for the same disease/condition in the past 12 months, and occurrence of minor, intermediate, and/or major complications. Residents were asked on the day of discharge if the patient could have been treated as an outpatient. Variables obtained from the discharge abstract were age, sex, race, home 50 miles or more from the hospital, discharge disposition, and the diagnosis-related groups (DRGs) and major diagnostic categories (MDCs).

ANALYTIC DATA SET

Data were collected for 901 admissions and over 10,000 patient days. None of the eligible patients and residents refused to participate. However, 73 cases were excluded, including 59 admissions still hospitalized at the end of the study; the 14 additional admissions not eligible included 10 transfers to non-study services and 4 with stays of less than 24 hours.

Among the 828 eligible cases, 661 (80 percent) had a complete set of checklists and a discharge abstract. The percentage of completed checklists was generally similar by resident. By virtue of their back-up status, the junior half of medical resident pairs completed lower percentages of the checklists.

The 1981 version of the DRG algorithm was used for the 661 admissions. Binary variables were constructed for the 13 DRGs that had at least ten admissions. All other DRGs were aggregated to the

level of their MDC, and binary variables were developed for each MDC.

STATISTICAL METHODS

Pearson correlations were used to test degree of overlap between the ordinal severity and difficulty measures, and Spearman correlations were used for those in binary form.

Hierarchical stepwise regression was used to test the explanatory power of the severity and difficulty measures for length of stay, controlling for other data.

The Kolmogorov-Smirnov One Sample Test was used to evaluate the normality of the length of stay distribution. Although not normal, the natural logarithm was the best transformation to permit linear regression while retaining outliers, which are an important source of variability.

RESULTS

Patient Characteristics. Table 1 displays characteristics of the 661 admissions analyzed. Almost 75 percent were 45 years of age or older, 68 percent were male, 66 percent were white, and about 80 percent lived within 50 miles of the hospital. Thirty-one percent had been hospitalized for the same problem during the previous year. Thirty-eight percent of the admissions were emergencies, approximately equally divided between medical and surgical cases. Regarding complications, 18.2 percent had complications "of any type," with 13.5 percent with at least one minor one during the stay, 7.1 percent at least one intermediate, and 3.5 percent at least one major complication. Ninety-four percent of these patients were discharged home, and 10 percent could have been treated as outpatients. The mean length of stay was 11.7 days (natural logarithm = 2.1).

Table 2 presents the diagnostic distribution. Twelve of a possible 21 MDCs are represented, and within these categories, frequencies of DRGs with at least ten admissions are shown. Two-thirds of the admissions are accounted for by five MDCs: circulatory for 17.7 percent, digestive for 17.2 percent, respiratory for 10 percent, musculoskeletal and connective tissue for 10.8 percent, and kidney for 8 percent. These MDCs, and the others in the data, span a wide range of adult medical and surgical conditions.

Table 1: Percentages of Patient Demographic and Admission and Discharge Characteristics

	Percent		*Percent*
N = 661*			
Age		Admission type	
<21	4.2	Medical	53.4
21-34	12.6	Surgical	46.6
35-44	9.4	Admissions with any complication	18.2
45-54	20.0		
55-64	29.3	Admissions with a minor complication at least once during the stay	13.5
65-74	13.9		
75+	10.6		
Sex		Admissions with an intermediate complication at least once during the stay	7.1
Male	67.8		
Female	32.2		
Race			
White	66.3	Admissions with a major complication at least once during the stay	3.5
Nonwhite	33.7		
Home ≥50 miles from hospital	18.9	Could have been treated as an outpatient	10.1
Hospitalized for same disease/condition in past 12 months	31.2	Discharge disposition	
		Home	94.4
Admission urgency		Other facility	2.7
Emergency	37.7	Against medical advice	1.4
Nonemergency	62.3	Death	1.5

*Mean of the natural logarithm of length of stay = 2.1.

Descriptive Statistics for Severity and Difficulty Measures. Table 3 presents means of the first-day, last-day, and peak severity measures, the daily average, and the percentage of admissions with fluctuating severity. Comparison of first- and last-day means reveals that, irrespective of length of stay, patients were rated more severe when they came in, less so at discharge. As length of stay increased, means of all measures increased; further, differences between first-day and peak values were greater, and the percentage of admissions with fluctuating severity increased.

Table 4 shows the means of first-day, last-day, and peak difficulty measures; the daily average; and percentages of admissions with fluctuating difficulty. As with severity, patients were consistently rated more difficult when they came in than when they went out and, with increased length of stay, the means of all measures increased, as did differences between first-day values and peaks and percentages of

Table 2: Percentages of Diagnosis-Related Groups and Major Diagnostic Categories

N = 661*		
MDC 1 Nervous System	Total‡	5.7
DRG 25	1.5	
Other MDC 1	4.2	
MDC 4 Respiratory System	Total	10.0
DRG 82 Respiratory neoplasms	1.5	
DRG 88 Chronic obstructive pulmonary disease	2.3	
Other MDC 4	6.2	
MDC 5 Circulatory System	Total	17.7
DRG 122 Circulatory disorders with acute myocardial infarction, no C.C.†	1.8	
DRG 127 Heart failure and shock	2.3	
DRG 131 Peripheral vascular disorders, age < 70, no C.C.	1.5	
DRG 140 Angina pectoris	2.1	
DRG 143 Chest pain	2.6	
Other MDC 5	7.4	
MDC 6 Digestive System	Total	17.2
DRG 162 Inguinal and femoral hernia procedures, age 18–69, no. C.C.	3.3	
DRG 183 Esophagitis, gastroenteritis, and miscellaneous digestive disease, age 18–69, no. C.C.	2.4	
Other MDC 6	11.5	
MDC 7 Hepatobiliary System and Pancreas	Total	3.9
MDC 8 Musculoskeletal System and Connective Tissue	Total	10.6
DRG 243 Medical back problems	3.8	
Other MDC 8	6.8	
MDC 9 Skin, Subcutaneous Tissue and Breast	Total	6.9
DRG 262 Breast biopsy and local excision for nonmalignancy	1.5	
DRG 284 Minor skin disorders, age < 70, no C.C.	1.5	
Other MDC 9	3.9	
MDC 10 Endocrine, Nutritional and Metabolic	Total	2.1
MDC 11 Kidney and urinary tract	Total	8.0
MDC 12 Male reproductive system	Total	4.2
MDC 13 Female reproductive system	Total	3.9
MDC 17 Myeloproliferative diseases	Total	1.8
DRG 468 Unrelated diagnosis and procedure	Total	7.7
		100.0

*DRGs with at least 10 cases are shown; all other DRGs are aggregated into MDCs.

†C.C. = Complications and Comorbidities.

‡Totals rounded.

Table 3: Means of Illness Severity Measures Stratified by Length of Stay

Length of Stay	*No. of Cases*	*Mean Values*				*Percent Fluctuating*
		Day 1	*Average*	*Peak*	*Last Day*	
Total	661	1.2	1.1	1.5	0.8	45.5
1-4 days	162	0.9	0.8	1.0	0.7	6.2
5-8 days	180	1.1	0.9	1.4	0.7	27.8
9-14 days	144	1.3	1.1	1.6	0.8	63.9
15-35 days	143	1.3	1.2	1.9	1.0	82.5
35+ days	32	1.6	1.6	2.5	1.4	96.9

Table 4: Means of Difficulty of Clinical Management Measures Stratified by Length of Stay

Length of Stay	*No. of Cases*	*Mean Values*				*Percent Fluctuating*
		Day 1	*Average*	*Peak*	*Last Day*	
Total	661	0.9	0.8	1.6	0.7	57.0
1-4 days	162	0.8	0.7	1.2	0.6	7.4
5-8 days	180	0.9	0.7	1.4	0.6	53.3
9-14 days	144	1.0	0.8	1.7	0.7	76.4
15-35 days	143	1.1	0.9	2.1	0.8	88.8
35+ days	32	1.4	1.2	2.6	1.0	100.0

admissions with fluctuating difficulty. In contrast, the first, last, and average values for difficulty were lower than those for severity; the peaks were slightly higher; and the percentages with fluctuating difficulty were much higher than those for severity reported in Table 3.

Table 5 displays Pearson correlations of the severity and difficulty measures. Generally, these are relatively high correlations, ranging from .53 to .74. Note that the peak and fluctuating measures are less correlated than the average measure, as are the first-day and last-day measures.

Table 6 presents hierarchical stepwise regression models for length of stay. MDCs and DRGs with at least ten admissions were entered in initially, followed by severity and by difficulty measures. Separate analyses were done for first-day, daily average, peak, and fluctuating measures. The DRGs/MDCs explained only 9 percent of length of stay. Adding severity measures substantially increased explanatory power, varying by the specific measure. Severity first-day and daily average measures explained 3 percent and 5 percent, respectively, while peak explained 21 percent, and fluctuating severity, 34 percent. Adding first-day difficulty increased explained variance by 1 percent,

Table 5: Pearson Correlations of Illness Severity and Difficulty of Clinical Management Measures

N = 661	*Measures Correlated*		
	Severity	*Difficulty*	*Correlation*
	Day 1	Day 1	.63
	Average	Average	.74
	Peak	Peak	.57
	Fluctuating	Fluctuating	.53*
	Last Day	Last Day	.70

*As the fluctuating measure is a zero-one variable, Spearman correlations also were tried. Results were similar.

Table 6: Hierarchical Regression Model for Natural Log Length of Stay, Coefficients and Explanatory Power of Illness Severity, Difficulty of Clinical Management, and Diagnosis Variables

	Day 1	*Average*	*Peak*	*Fluctuating*
N = 661				
Step 1				
DRG 143	-.64	-.57	-.47	-.45
DRG 262	-.78	-.73	-.54	-.21*
DRG 82	.78	.61	.42	.34
MDC 4	.31	.32	.19*	.25
MDC 6	.24	.24	.14*	.12*
DRG 122	.42*	.47	.29*	.14*
MDC 5	.17*	.19*	.05*	.14*
MDC 7	.29*	.25*	.06*	.12*
Step 2				
Severity	.11	.30	.35	.63
Step 3				
Difficulty	.12	—†	.23	.61
Intercept	1.83	1.76	1.21	1.46
R^2, 1	.09	.09	.09	.09
R^2, 1+2	.12	.14	.30	.43
R^2, 1+2+3	.13	.14	.34	.53

*With addition of severity and difficulty, asterisked variables became statistically not significant ($p > .05$).

†Did not enter.

while the daily average added 0 percent, and the peak and fluctuating difficulty measures added 4 percent and 10 percent, respectively. These results measured the independent value of difficulty as a predictor for length of stay once severity is controlled for. They also served to

suggest the importance of using a particular choice of measure. The peak and fluctuating measures performed better than the first-day and average measures.

DISCUSSION

In this study, physicians directly responsible for managing acute inpatient care completed daily ratings for each of their patients to assess severity of illness and the difficulty of the patient's clinical management. Ratings were used to determine the feasibility of a scale to measure difficulty of clinical management, and whether or not difficulty was independent of severity.

As expected, not only is difficulty related to severity; it also adds a dimension to patient care that aids in explaining variations in length of stay. Once DRG/MDC categories plus severity are taken into account, difficulty adds up to 10 percent of the variance explained. These results indicate that, to the extent that severity is used as a surrogate for the difficulty of clinical management, it is a good but imperfect proxy. It is important to appreciate that it is not severity alone that affects variations in length of stay within DRGs/MDCs.

Four measures of severity and difficulty were compared in this study: first-day, average, peak value, and whether or not the pattern fluctuated over the stay. Neither first-day nor average were highly predictive of length of stay, but peak values and fluctuating patterns were strongly predictive.

The use of peak values is not new. Horn's severity measures use peak values (Iezzoni, Moskowitz, and Daley 1989: Horn, Sharkey, Buckle, et al. 1991) as do the MedisGroups (Iezzoni, Ash, and Moskowitz 1987) and APACHE II and III measures (Wagner, Draper, and Knaus 1989; Knaus, Wagner, Draper, et al. 1991). In contrast, fluctuating patterns is a previously untested concept. To be classified as having a fluctuating pattern, a patient must experience a level of severity (difficulty) during the stay that is above (below) both the first-day and last-day values. Literally, this means that during the hospital stay the patient either increases in severity or difficulty, followed by a decline, or decreases in severity or difficulty, followed by an increase. Thus, it is not surprising that patients reported to have had complications are more likely than those without complications to experience fluctuating severity (78.3 percent versus 38.3 percent) or fluctuating difficulty (85 percent versus 50.8 percent). Clearly, however, the

occurrence of complications does not fully explain fluctuating patterns.

This study represents an unusual commitment by residents in collecting patient-specific data for research. Over 10,000 patient days of care were rated to provide the data for analyzing patterns of severity and difficulty of clinical management. All patients who were approached to participate in the study agreed to do so. Only 8 percent of the patient-day forms were not completed. Even so, there are several significant limitations to this study. First, there was not a test of interrater reliability in scoring patients each day. It would have been desirable to know the reliability of ratings by different residents, and between residents and attending physicians. A related shortcoming is that there were only a small number of residents, particularly in surgical specialties, and not all training levels were well represented.

Second, the patient population and the facility within which the study was done are not representative. Even so, the types of conditions treated, as reflected by the DRG/MDC distribution, include many commonly seen among adult medical/surgical patients in general acute care hospitals. Only 10 percent of patients were rated as potentially treatable on an outpatient basis, very comparable to the percentage thus rated in general hospitals before the introduction of managed care criteria used by most insurers today.

It is unknown how DRG-based prospective payment and the associated shorter hospital stays would affect the findings presented. It is expected that the findings would be similar, although evaluation of the Medicare prospective payment system indicates that the elderly are being discharged "sicker" under PPS than before its introduction (Kosecoff, Kahn, Rogers, et al. 1990), and that, for selected conditions, they are sicker at admission (Keeler, Kahn, Draper, et al. 1990). This might affect first-day and average severity and the fluctuating measures modestly, but would be unlikely to affect the peak measures.

In summary, the difficulty of clinical management can be measured. It is related to severity, yet provides additional information that aids in explaining the use of inpatient resources as measured by length of stay. Unfortunately, information was unavailable on charges or costs of care at the individual patient level.

Although the findings of this one study cannot be broadly generalized, the results strongly suggest the value of capturing information relevant for classifying patients in terms of severity and difficulty. Physician reporting may not be the most desirable or practical way to do this. Adaptation of successful chart-based methods for scoring severity is worth considering as a strategy for scoring difficulty of

clinical management. Whether or not the cost of obtaining this additional information would be worth the investment is still an open question for the routine measurement of severity, as it is likely to be for measuring the difficulty of clinical management.

ACKNOWLEDGMENTS

This study was part of the author's doctoral dissertation for the Department of Health Policy and Management, the Johns Hopkins University School of Hygiene and Public Health. The author thanks her committee, including Alice Gifford, Alan Gittelsohn, David Levine, and David Salkever, as well as ex officio member Laura Morlock, and Charles Flagle, who introduced her to the individuals at the Baltimore U.S. Public Health Service Hospital who, in turn, made it possible to do the study; these include David Ross, Patrick Mattingly, and Stephen Jencks. Special thanks also go to the study physicians; to Bart Bennet, Joel Gibson, Seth Perlman, and Paul Novotny, who were computer programmers at various times; to Terri Bhat, who did the word processing for the manuscript; and to Donald Steinwachs, who gave invaluable help in revising this article.

REFERENCES

Aronow, D. B. "Severity of Illness Measurement: Applications in Quality Assurance and Utilization Review." *Medical Care Reveiw* 45, no. 2 (Fall 1988): 339–66.

Bohrnstedt, G. W. "A Quick Method for Determining the Reliability and Validity of Multiple-Item Scales." *American Sociological Review* 34, no. 4 (August 1969): 542–48.

Burns, L. R., and D. R. Wholey. "The Effects of Patient, Hospital, and Physician Characteristics on Length of Stay and Mortality." *Medical Care* 29, no. 3 (March 1991): 251–71.

Bursztajn, H., R. I. Feinbloom, R. M. Hamm, and A. Brodsky. *Medical Choices, Medical Chances. How Patients, Families, and Physicians Can Cope with Uncertainty.* New York: Delacorte Press/Lawrence, 1981.

Calore, K. A., and L. Iezzoni. "Disease Staging and PMCs: Can They Improve DRGs?" *Medical Care* 25, no. 8 (August 1987): 724–37.

Cretin, S., and L. G. Worthman. *Alternative Systems for Case Mix Classification in Health Care Financing.* Publication no. R-3457-HCFA. Santa Monica, CA: RAND Corporation, 1986.

Cronbach, L. J. "Co-efficient Alpha and Internal Structure of Tests." *Psychometrika* 16, no. 3 (1951): 297–334.

Engelhardt, H. T., S. F. Spicker, and B. Towers, eds. *Clinical Judgment: A*

Critical Appraisal. Proceedings of the Fifth Trans-Disciplinary Symposium of Philosophy and Medicine, Vol. 6. Los Angeles, CA, April 14–16, 1977. Dordrecht, Holland: D. Reidel Publishing Co., 1979.

Feinstein, A. R. *Clinical Judgment.* Baltimore, MD: Williams & Wilkins, 1967.

Fox, R. C. "The Evolution of Medical Uncertainty." *Milbank Memorial Fund Quarterly* 58, no. 1 (Winter 1980): 1–49.

Horn, S. D., P. D. Sharkey, J. M. Buckle, J. E. Backofen, R. F. Averill, and R. A. Horn. "The Relationship between Severity of Illness and Hospital Length of Stay and Mortality." *Medical Care* 29, no. 4 (April 1991): 305–17.

Iezzoni, L. I., A. S. Ash, and M. A. Moskowitz. "MedisGroups: A Clinical and Analytic Assessment." Health Policy Research Consortium. Prepared for the Health Care Financing Administration under Cooperative Agreement no. 18-C-98526/1-03. Baltimore, MD, July 1987.

Iezzoni, L. I., M. A. Moskowitz, and J. Daley. "A Description and Clinical Assessment of the Computerized Severity Index." *Health Policy Research Consortium.* Prepared for the Health Care Financing Administration under Cooperative Agreement no. 18-C-99526/1-05. Baltimore, MD, July 1989.

Iezzoni, L. I., M. Schwartz, and J. Restuccia. "The Role of Severity Information in Health Policy Debates: A Survey of State and Regional Concerns." *Inquiry* 28 (Summer 1991): 117–28.

Jencks, S. F., and A. Dobson. "Refining Case-Mix Adjustment: The Research Evidence." *New England Journal of Medicine* 317, no. 11 (September 1987): 679–86.

Keeler, E. B., K. L. Kahn, D. Draper, M. J. Sherwood, L. V. Rubenstein, E. J. Reinisch, J. Kosecoff, and R. H. Brook. "Changes in Sickness at Admission following the Introduction of the Prospective Payment System." *Journal of the American Medical Association* 264, no. 15 (October 1990): 1962–68.

Knaus, W. A., D. P. Wagner, E. A. Draper, J. E. Zimmerman, M. Bergner, P. G. Bastos, C. A. Sirio, D. J. Murphy, T. Lotring, A. Damiano, and F. E. Harrell. "The APACHE III Prognostic System. Risk Prediction of Hospital Mortality for Critically Ill Hospitalized Adults." *Chest* 100 (December 1991): 619–36.

Kosecoff, J., K. L. Kahn, W. H. Rogers, E. J. Reinisch, M. J. Sherwood, L. V. Rubenstein, D. Draper, C. P. Roth, C. Chew, and R. H. Brook. "Prospective Payment System and Impairment at Discharge. The 'Quicker-and-Sicker' Story Revisited." *Journal of the American Medical Association* 264, no. 15 (October 1990): 1980–83.

Mackenzie, T., A. R. Willan, and the Case Mix Research Group. *Patient Classification Systems: An Evaluation of the State of the Art.* Vol. 1. Prepared for the Health Care Financing Administration under Agreement no. 17-C-99133/F-02. Kingston, Ontario: Queens University, February 1991.

McMahon, L. F., and J. E. Billi. "Measurement of Severity of Illness and the Medicare Prospective Payment System: State of the Art and Future Directions." *Journal of General Internal Medicine* 3, no. 5 (September/October 1988): 482–90.

Rosko, M. D. "DRGs and Severity of Illness Measures: An Analysis of

Patient Classification Systems." *Journal of Medical Systems* 12, no. 4 (August 1988): 257–74.

Schoonhoven, C. B., W. R. Scott, A. B. Flood, and W. H. Forrest. "Measuring the Complexity and Uncertainty of Surgery and Post Surgical Care." *Medical Care* 18, no. 9 (September 1980): 893–915.

Thomas, J. W., M. Ashcraft, and J. Zimmerman. *An Evaluation of Alternative Severity of Illness Measures for Use by University Hospitals.* Vols. 1–3. Ann Arbor, MI: Department of Health Services Management and Policy, School of Public Health, The University of Michigan, 1986.

Wagner, D. P., E. A. Draper, and W. A. Knaus. "Development of APACHE III." *Critical Care Medicine* 17, no. 12, part 2/supplement (December 1989): S199–S203.