

Profiles of Physician Practice and Patient Severity of Illness

SUSAN D. HORN, PHD, ROGER A. HORN, PHD, AND HAMILTON MOSES, MD

Abstract: We report on a study that examined physician practice profiles using two methods of patient classification: the Severity of Illness Index and diagnosis-related groups (DRGs). When used together with conventional management information and DRGs, the Severity of Illness Index permitted useful comparisons to be made among physicians; differences in both case-mix and severity could be estimated. In 37 per cent of the physicians studied, we found differences of more than \$10,000 in the apparent impact of a physician

on the hospital's financial position, depending on whether one controlled for severity or not. The extent to which these differences in impact could be due to quality of care differences is an area for future research. However, the findings that 37 per cent of the physicians in the study may be wrongly identified as over- or under-utilizers suggest long-term public health consequences of preparing physician profiles based on unadjusted DRGs. (*Am J Public Health* 1986; 76:532-535.)

Introduction

Physicians strongly influence hospital health care costs. Thus indicators of physician practice patterns and their impact on resource consumption are important to efforts to reduce health care costs.

Many case mix grouping systems and measures of patient characteristics have been used to describe patient populations and their consumption of resources. Experience has shown that management statistics, such as total dollars expended per patient, or the average number of procedures performed, or the average length of hospital stay, are inadequate as useful comparisons among physicians. These statistics do not account for differences in the severity of illness of each physician's patients, which often affect much of the inter-physician variability in management statistics. Even systems that attempt to adjust for case mix, such as the diagnosis-related groups (DRGs), have not always been effective; they do not control for severity of illness.

Eisenberg and Williams¹ reviewed several strategies to contain costs by changing physicians' practice and behavior. They summarized these efforts in several categories, including educational efforts and administrative changes (especially in the use of routine and pro forma tests) as well as economic incentives. None of these strategies has been uniformly successful. Some have had value for a short time, but behavior reverted to its former state soon after the experiments ended. Thus, the observed changes may have been only a medical Hawthorn effect. All of these strategies have the shortcoming that the information used for comparisons and the determination of standards of care are based on management statistics that fail to account for differences in patient severity of illness or the presence of other interacting factors.

We report on a study of the use of the Severity of Illness Index² to provide a more rational method for inter-physician comparisons. When used together with conventional management information and DRGs, the Severity of Illness Index allows useful comparisons to be made among physicians because differences in both casemix and severity can be estimated, and one can analyze both economic data and clinically valid information on an ongoing basis.

Address reprint requests to Susan D. Horn, PhD, Associate Director, Center for Hospital Finance and Management, Department of Health Policy and Management, Johns Hopkins University, 624 N. Broadway, Baltimore, MD 21205. Dr. R. Horn is Professor, Department of Mathematical Sciences, and Dr. Moses is Deputy Chairman, Department of Neurology, School of Medicine, all at JHU. This paper, submitted to the *Journal* June 27, 1985, was revised and accepted for publication September 11, 1985.

© 1986 American Journal of Public Health 0090-0036/86\$1.50

Methods

We studied patients treated during the six-month period July 1, 1984 through December 31, 1984 in a university teaching hospital. There were 14,585 patients discharged in this period. We obtained complete discharge abstract and financial data for each patient, including principal diagnosis, up to four secondary diagnoses, up to six procedures, patient disposition, age, payer status, attending physician, and charges. These data were used to classify the patients into a total of 456 (of the 467) DRG categories representing all of the 23 major diagnostic categories (MDCs).

To determine the patient's severity of illness level, each patient's medical record was examined after discharge by one of 16 nurses who were trained in the use of the Severity of Illness Index. The Severity of Illness Index, described in detail elsewhere,³ takes into account seven dimensions about the patient:

- 1) Stage of principal diagnosis,
- 2) Complications of the principal diagnosis,
- 3) Concurrent interacting conditions that affect the hospital course,
- 4) Patient dependency on hospital facilities and staff,
- 5) Extent of non-operating room life support procedures,
- 6) Rate of response to therapy or rate of recovery, and
- 7) Impairment remaining after therapy for the acute aspect of the hospitalization.

The seven dimensions are used as a guide to help the rater score the overall severity of illness of the patient from Level 1 (least severe) to Level 4 (most severe).

Quality of care is not addressed directly by the Severity of Illness Index or by the DRG system. A patient may be assigned to a higher severity of illness level because of either natural or iatrogenic factors. Better quality of care may result in lower severity of illness levels for some patients. The Severity of Illness Index therefore quantifies only the de facto severity of the patient's illness, independent of the cause of the severity. However, methods are being developed to use it as a screen to flag possible quality of care problems.

The reliability of the nurse raters in scoring severity of illness was checked monthly for the first three months and again at six months. Overall, each nurse rater achieved at least 90 per cent agreement when samples of cases were re-rated by independent experienced raters. The average agreement was 95.1 per cent.

The final combined data set consisting of discharge abstract data, severity of illness data, and financial data was subjected to a computer edit to identify anomalous cases before the data were analyzed. About 0.5 per cent of the cases

were flagged by the editing program as anomalous. These cases included missing data, length of stay, and charge anomalies (such as zero length of stay or charges per day less than the hospital's routine daily charge, which are apparent errors in hospital data), and severity of illness pattern anomalies (such as all dimensions at Level 1 but overall severity of illness at Level 4, which is a probable key punch error). The hospital was requested to review these cases and to correct errors before the data were analyzed.

A computer algorithm was used to assign each case to one of three procedure type codes: no operating room (OR) procedure, moderate OR procedure, or major OR procedure. The master list of procedure types was prepared by surgeon consultants who examined every ICD-9-CM procedure code. Major operating room procedures are those that either take special skills and extensive training to perform or take a long time to recover from, or both. About 15 per cent of all operating room procedures are classified as *major* and include such operating room procedures as craniotomy, open heart surgery, total hip replacement, laminectomy, etc. All other operating room procedures are classified as *moderate*.

We grouped the study patients into four primary sets of groups:

- 1) DRGs (DRG),
- 2) Severity of illness and procedure type within MDCs (MDC/Sev/Pr),
- 3) Severity of illness within DRGs (DRG/Sev), and
- 4) Severity of illness and procedure type within DRGs (DRG/Sev/Pr).

Two additional sets of groups were produced by subdividing according to attending physician within the primary sets of groups 1 and 4. We examined the homogeneity of total charges for the groups produced in these six ways using three statistical measures of homogeneity⁴:

- a) Reduction in variance (RIV),
- b) Coefficient of variation (CV), and
- c) Analysis of variance F test.

Because the DRG/Sev/Pr set of groups was the most homogeneous of the four primary sets of groups, we computed the arithmetic average resource use within these groups as quantified by total charges, laboratory charges, radiology charges, pharmacy charges, routine charges, and length of stay; these averages became the norms against which to compare each physician's patients' resource use. For each of the patients treated by a physician, we compared that patient's resource use in each resource use category to the norms for the respective DRG/Sev/Pr groups and accumulated the differences between actual resource use and the respective norms across all patients treated by that physician. A positive (+) accumulated difference indicates that the physician was using more resources compared with the norms based on DRG/Sev/Pr. A negative (-) accumulated difference indicates that the physician was using less resources than the norms based on DRG/Sev/Pr. We repeated this same analysis, but used DRGs only (not adjusted for severity of illness) to produce norms against which to compare each patient and subsequently each physician's patients. This was done to determine if the financial impact of a physician's practice looked different when DRGs were used to classify the patients rather than Sev/Pr-adjusted DRGs. It has been stated that even though DRGs may not describe each individual patient's resource use well, they are useful to describe the aggregated resource needs of a hospital or a physician's practice.⁵

TABLE 1—Charge Homogeneity Statistics by Case Mix Grouping System Sample Size (N = 14,585 Patients)

Method	No. Groups	RIV	CV	F
DRG	456	44	62	24
Sev/Pr	12	53	58	1364
MDC/Sev/Pr	236	62	49	100
DRG/Sev	1174	71	42	28
DRG/Sev/Pr	1416	75	41	28

TABLE 2—Effects of Physicians On Case Mix Grouping Systems Sample Size (N = 14,585 Patients)

Method	No. Groups	RIV	CV	F
DRG	456	44	62	24
DRG/MD	5,046	75	114	6
DRG/Sev/Pr	1,416	75	41	28
DRG/Sev/Pr/MD	6,523	92	30	14

TABLE 3—Number of Physicians with Positive or Negative Impact on Hospital Financial Status Using DRG or DRG/Sev/Pr Norms

	DRG/Sev/Pr Impact		
	+	-	
DRG Impact	+	111	36
	-	85	186
			418

- = using less resource use than norms.
 + = using more resource use than norms.

Results

We examined the homogeneity of all four primary sets of groups. Even though the MDC/Sev/Pr methodology resulted in fewer case mix groups than the DRG methodology, the set of MDC/Sev/Pr groups was more homogeneous than the set of DRG groups as indicated by much higher RIV and F values and lower CV. Table 1 shows the homogeneity statistics for these data. The DRG/Sev/Pr groups have the greatest homogeneity.

In Table 2 we show what happens in DRGs and DRG/Sev/Pr when the data are further grouped according to attending physician. DRGs alone explain 44 per cent of the variability in resource use. When DRGs are subdivided by attending physician, 75 per cent of the variability is explained, but the number of groups increases ten-fold and the weighted average coefficient of variation rises dramatically, indicating that the groups containing more than one case are very heterogeneous; groups containing only one case have zero CV and hence do not enter into the computation of the weighted CV. On the other hand, when DRGs are divided into severity and procedure subgroups, 75 per cent of the variability is explained again, but with thousands of fewer groups and with much smaller weighted CV and much higher F values, indicating that these groups are more homogeneous than the DRG/MD groups. When the DRG/Sev/Pr groups are further subdivided by attending physician, 92 per cent of the variability in resource use is explained and the weighted CV is the smallest among all the analyses.

In Table 3 we show a summary of the financial impact of physician profiles when DRGs are used as the standard to define patients compared with DRG/Sev/Pr groups. We see

TABLE 4—Absolute Change in Physicians' Financial Impact from DRG Norms to Sev/Pr-Adjusted DRG Norms

Absolute Change Dollar Impact (\$)	Number of Physicians
0-100	20
100-1,000	71
1,001-5,000	116
5,001-10,000	58
10,001-15,000	41
15,001-20,000	24
20,001-30,000	28
30,001-40,000	15
40,001-50,000	11
50,001-75,000	14
75,001-100,000	7
>100,001	13

TABLE 5—153 Physicians with Absolute Change in Impact Greater than \$10,000 by DRG vs DRG/Sev/Pr Norms

DRG	DRG/Sev/Pr		
	+	-	
+	39	23	153
-	26	65	

- = using less resource use than norms.
 + = using more resource use than norms.

that 71 per cent (297) of the 418 physicians in this study had financial impact of the same sign under DRGs and DRG/Sev/Pr groups; the financial status of these physicians as + (using more resources than the norms) or - (using less resource use than the norms) was the same under both the DRGs and the more refined Sev/Pr-adjusted DRGs. The financial impact of 29 per cent (121/418) of the physicians on the hospital was of a different sign under the two case mix systems. For 36 physicians, there was more resource use (+) than the DRG norms but less resource use (-) when the norms were adjusted for Sev/Pr. For 85 physicians, there was less resource use than the norms (-) under DRGs but more resource use (+) when the norms were adjusted for Sev/Pr.

The difference in signs (+ or -) of a physician's practice under DRGs or DRG/Sev/Pr groups tells only part of the story. One should also look at the magnitude of the difference in financial impact under these two case mix systems. In Table 4 we show the summary distribution of the absolute change of the magnitudes (in dollars) of the overall financial impact of each physician from DRG norms to Sev/Pr adjusted DRG norms.

We discussed these distribution results with several hospital administrators and physicians to determine the magnitude of the absolute change that they considered important to detect. They felt that \$10,000 or more absolute change was meaningful to them. If we consider an absolute change of less than \$10,000 using norms based on the two case mix systems as meaning that the two systems provide roughly the same picture of the impact of a physician's practice on a hospital's financial status, then we find 63 per cent (265/418) of the physicians met this criterion and 37 per cent (153/418) did not. The 153 physicians whose difference was greater than \$10,000 had the distribution of signs of impact shown in Table 5.

We found that 42 per cent (65/153) of the physicians

whose difference in impact was greater than \$10,000 under the two systems of case mix norms had less resource use (-) than the norms under both case mix systems. Another 25 per cent (39/153) had more resource use (+) than the norms under both case mix systems. Thus, a total of 68 per cent (104/153) of the physicians whose difference in impact was greater than \$10,000 under the two case mix system norms had the same conclusions about the sign of their impact, but a different conclusion about the magnitude of their impact. The remaining 32 per cent (49/153) of the physicians differed with respect to both sign and magnitude. Expressed as a per cent of the total sample, this was 12 per cent (49/418) of the physicians.

For management purposes within the hospital, we also displayed physician-specific detailed comparisons. Table 6 shows an example of the detailed results for one physician and one DRG. For each patient and each category of resource use, no sign (an implicit "+") preceding the amount indicates that the patient consumed more resources than the norm in that resource category. Recall that the norm in a resource category is the arithmetic average for all patients in the same DRG, severity level, and procedure type in the whole data set. A "-" sign preceding the amount indicates less resource use than the norm. Many of physician A's patients in DRG 148 have "-" signs for the various categories of resource use except for the patients in severity level 4, both of whom lived. The same is true for physician A's patients in other DRGs (65 per cent of his patients have charges less than the norm). Table 6 also shows how ancillary charges and length of stay differ from their respective norms.

To summarize each physician's effect on the hospital's financial picture, we summed the differences between actual patient charges and the norm charges over all of the physician's patients. Table 7 shows these results for several physicians as well as the results when their cases are grouped by DRGs only, without any adjustment for severity of illness or procedure type. All types of differences are observed. For the overall data, the results using the two case mix systems are very similar for physician D. Thus, adjusting for severity of illness and procedure type does not make much difference for this physician. However, for physicians A, B, and C the results are very different. Unadjusted for severity of illness and procedure type, physicians A and C look as though they use more resources than the norm, but when we control for the severity and procedure type, we come to the opposite conclusion for physician A and we find that physician C is closer to the norms (zero), although still positive. Physician B shows the same sign of impact but the magnitude is very different under the two case mix systems.

Discussion

In this study, we have demonstrated the difference in explanatory power made by incorporating a measure of severity of illness when examining the financial impact of a physician's practice on a hospital. Nearly all strategies of cost containment have been unpopular with physicians. In part, the objections have been due to apparent logical deficiencies in the ways data on physicians and patients have been collected and used. A nearly universal objection to these strategies has been a physician's complaint that "my patients are different, and your system doesn't account for it." It is natural for an apparently high-cost physician to feel that his patients are more severely ill, have more co-existing conditions, or respond slower to therapy; he may also feel that the quality of his medical practice is higher. None of these arguments is easy to evaluate with conventional data.

TABLE 6—Deviations from DRG/Sev/Pr Norms*

Physician	DRG	Sev	Pr	N	Total Charges	LOS (Days)	Rtn Charges	Rad Charges	Lab Charges	Pharm Charges
A	148	2	1	1	-359	0	-542	561	50	298
A	148	3	1	1	501	-3	51	318	420	-546
A	148	1	3	1	-1,094	-1	-260	-187	-28	-207
A	148	2	3	6	-1,451	-3	-1,021	20	21	1,081
A	148	3	3	3	-314	6	-1,479	-446	-647	-14
A	148	4	3	2	14,521	22	6,262	1,546	783	2,863
Overall A:				110	-7,027	13	-13,612	274	-1,604	-4,155

*The entries in the six columns from Total Charges through Pharmacy Charges represent combined data from the number of patients indicated in the column labeled N.
 LOS = Length of Stay.
 Rtn Charges = Routine Charges.
 Rad Charges = Radiology Charges.
 Lab Charges = Laboratory Charges.
 Pharm Charges = Pharmacy Charges.

TABLE 7—Deviations from DRG/Sev/Pr Norms and DRG Norms*

Physicians	N	Total Charges	LOS (Days)	Rtn Charges	Rad Charges	Lab Charges	Pharm Charges
A							
DRG/Sev/Pr	110	-7,027	13	-13,612	274	-1,604	-4,155
DRG	110	84,287	178	27,090	10,649	7,792	6,096
B							
DRG/Sev/Pr	328	-12,831	1	-38,498	-5,049	17,124	26,874
DRG	328	-71,057	1	-63,693	-8,112	7,061	23,979
C							
DRG/Sev/Pr	396	7,207	-16	-7,905	-3,500	-4,830	808
DRG	396	55,821	92	16,269	-2,345	-4,205	11,309
D							
DRG/Sev/Pr	66	-7,950	41	21,801	-3,648	-5,336	-4,460
DRG	66	-7,920	44	26,042	-6,085	-5,461	-4,911

*The entries in each of the columns from Total Charges to Pharmacy Charges represent the combined data from the number of patients indicated in the column labeled N.
 LOS = Length of Stay.
 Rtn Charges = Routine Charges.
 Rad Charges = Radiology Charges.
 Lab Charges = Laboratory Charges.
 Pharm Charges = Pharmacy Charges.

Information that is based solely on expense or charges, without adjustment for severity of illness, makes clinical comparisons difficult and is unlikely to have the intellectual support of the physicians whose practice patterns one is trying to affect. The Severity of Illness Index, however, provides a mechanism to describe these clinical factors in a reproducible, standard, and easily validated manner and hence allows useful comparisons to be made among physicians.

One limitation of this study is that the Severity of Illness Index, in its present form, does not account directly for differences in quality of care. Although the results discussed above are important in that they remove severity differences from the comparisons of physician financial impact, a final interpretation of individual physicians' deviations requires case-by-case consideration by peers.

If one wishes to implement a system to change hospital physicians' practice patterns, we believe that four principles should be kept in mind:

- Physicians must be approached with clearly defined goals for optimal practice patterns; often protocols can be defined.
- A strategy to accomplish the desired change must be clearly formulated.

- The method used to characterize practice patterns must be valid and clinically acceptable to the physicians.

- Some mechanism must exist to measure changes in practice patterns and signal when the goals have been achieved.

We believe that a valid measure of severity of illness is an essential component of any such system.

ACKNOWLEDGMENT

This research supported by the Pew Memorial Trust.

REFERENCES

1. Eisenberg JM, Williams SV: Cost containment and changing physicians practice behavior. *JAMA* 1981; 246:2195-2201.
2. Horn SD, Bulkley G, Sharkey PD, Chambers AF, Horn RA, Schramm CJ: Inter-hospital differences in patient severity: problems for prospective payment based on diagnosis related groups (DRGs). *N Engl J Med* 1985; 313:20-24.
3. Horn SD, Horn RA: Reliability and validity of the Severity of Illness Index. *Medical Care* (in press).
4. Horn SD, Horn RA, Sharkey PD, Chambers AF: Severity of illness within DRGs: homogeneity study. *Medical Care* (in press).
5. Jencks SF, Goldman HH, McGuire TG: Challenges in bringing exempt psychiatric services under a prospective payment system. *Hosp Community Psychiatry* 1985; 36:764-769.