7 .

Summary and Implications for Research and Policy

7

Summary and Implications for Research and Policy

SUMMARY

This case study has demonstrated that large dif - ferences exist in hospital length of stay (LOS) among geographic regions in the United States. The Northeast region has consistently exhibited the longest lengths of stay in the Nation and the West the shortest. Since 1968, despite a trend toward shorter LOS across all regions, the differences between East and West have persisted at about the same level. Although no attempt was made here to place a dollar value on these differences, there can be no doubt that the magnitude of the variation in LOS is potentially of considerable economic significance. The evidence is persuasive that these differences are not a function of differences in population characteristics such as age, sex, or race.

The extent to which case mix or severity of illness differences can account for LOS variations is more controversial. Adjusting regional LOS for differences in the distribution of diagnostic groups at a general level of aggregation reveals little effect on regional differences in LOS. However, severity of illness could explain some of the observed variation in regional LOS. This could occur if physicians in the Northeast systematically admit to the hospital only the sickest patients in each diagnostic category while those in the West admit the least sick patients.

Case mix differences among regions have rarely been studied at this level of clinical detail. In the Professional Standards Review Organization (PSRO) studies discussed in chapter 2, five conditions were studied in such a way as to be able to draw inferences concerning differences in severity of illness. In the Utah/Central Massachusetts studies there was a small, but statistically significant difference in seventy of illness of cholecystectomy patients, with the eastern PSRO experiencing the more severe case mix. There was no difference in case mix for myocardial infarction (MI) patients. In the Baltimore/Portland studies, there were significant differences in case mix for MI and congestive heart failure (CHF) patients but not for angina patients. In both cases where the distribution of cases was different, the eastern PSRO experienced the more complex case mix.

In two of the three examples of differing case mix, it was possible to adjust average LOS for these differences. Table 13 presents these results.

Table 13.—Case Mix Differences Between Baltimore and Portland

	Bait i more		Portland	
Condition and severity class	Number of cases	LOS (days)	Number of cases	LOS (days)
1. Uncomplicated MI	21	13.6	11	8.0
2. Moderately complicated MI	88	16,9	63	11.4
3. Severely complicated MI	26	20.5	4	9.3
Total	135	17.1	78	10,8
Adjusted MI LOS		(16.9)		(10.6)
1. Uncomplicated CHF	15	9.1	13	4 7
2. Moderately complicated CHF	27	11,3	34	8.3
3. Severely complicated CHF	71	14.0	19	7.0
Total	113	12.7	66	7,2
Adjusted CHF LOS		(12.3)		(7,1)

Calculated from data in A Ankrum Baltimore City Professional Standards Review Organ ization, personal communication November 1982

The adjustments were performed by using the entire study population of both PSROs as the reference population and then by applying each PSROs severity-class-specific LOS to the reference population. Thus, adjusting for case mix differences failed to affect the difference between PSROs in average LOS for MI patients and reduced the difference for CHF patients by only 5 percent. This might be explained by the fact that while LOS was related directly to severity of illness in the Baltimore patient population, the Portland patients did not show the same simple relationship. In Portland, the most severe patient group in both cases had a shorter LOS than the group with moderate severity. This, in turn, may have occurred because there were more early in-hospital deaths in the Portland group of most severely ill patients. These data are not contained in the reports from which this information is drawn. This apparent discrepancy may also be due to a failure to adequately define severity of illness.

Whatever the reason for the short LOS in the Portland patient populations, it is apparent that adjusting for case mix differences in these two instances does not diminish differences in average LOS. On the other hand, it is also true that for these two conditions and for each of the other three studied, LOS was shorter in the western PSRO in each severity class. In reviewing all of the evidence, therefore, while allowing the possibility that some case mix differences between East and West may exist, one must conclude that such differences are most unlikely to account for a large part of observed regional LOS variations.

If substantial differences in LOS remain after controlling for case mix, one is forced to conclude that physicians must employ different treatments for the same conditions in the East and West. Research in this crucial area simply does not exist. The Utah/Central Massachusetts PSRO study provided some intriguing data on differing practices with respect to ambulation and feeding of postoperative cholecystectomy patients. Except for this example, there has been no attempt to gather this kind of information at a detailed clinical level. Furthermore, no studies have adequately addressed the question of whether these different lengths of stay have any impact on health status. There is simply no evidence as to whether western patients fare better or worse with their short lengths of stay compared to eastern patients.

In order to address the critical question of how hospital LOS is related to health outcome, the medical literature has been reviewed to discover the extent to which good quality research had established what LOS produces the best health outcome in specific clinical conditions. The goal of this analysis was to ascertain whether LOS standards could be inferred from these research studies and used to assess the appropriateness of regional LOS differences. If a medically optimal LOS could be determined from analyzing the medical literature, then regional LOS patterns could be compared to this standard, evaluating which region's LOS is too high and which is too low.

The medical literature has been examined, focusing on studies in which researchers attempted to change LOS for specific medical conditions in order to improve health outcomes. Randomized clinical trials (RCTs) have been given special attention in this review of the medical literature, because their design is most likely to produce valid results. Study and control groups are as comparable as possible in order to be as certain as possible that observed differences in outcomes are attributable to the experimental treatment. The RCT is certainly not a guarantee of such results and generalizability may be limited. Nor is the RCT the only informative design for medical research. Nevertheless, it is the most effective approach to the complex questions addressed here.

At least one methodologically sound RCT was found in five different clinical areas: MI, elective surgery (primarily inguinal herniorrhaphy), low-risk obstetrics, low birth weight infants, and psychiatry. All of these RCTs experimented with shorter lengths of stay than had been traditionally used, none with longer. All of the studies concluded that the shorter LOS was not harmful. Table 14 summarizes the overall characteristics and results of these 17 RCTs. In reviewing these studies, it is noteworthy that all but one study in each of the MI and surgery categories excluded the elderly from participation. Therefore, even the limited conclusions one can draw from these studies do not apply to the elderly. In addition,

virtually all of these studies excluded a significant proportion of patients screened for potential participation, the single exception being the study on low birth weight infants from Memphis. In all cases, patients were excluded because they were felt to be too sick to be candidates for early discharge.

Only two studies discussed mortality as an outcome: MI and low birth weight infants. The trend in the MI studies was for the groups with the shorter LOS to have slightly lower mortality, but these differences did not achieve statistical significance at the 5-percent level. There were no differences between groups in the low birth weight infant studies; the two studies combined reported only a single death in each patient group.

In table 14, morbidity is defined as follows: for MI patients, the rate of nonfatal cardiovascular complications during the followup period; for elective surgery, the rate of postoperative complications; for low-risk obstetrics, the rate of neonatal complications; and for low birth weight infants and psychiatry, the rate of readmission during the followup period, Only one surgery study and three psychiatry studies reported that their differences in morbidity were significant at the 5-percent level. However, the authors of the surgical studies stated uniformly that the complications observed were of little clinical significance and should not be considered reason enough to discontinue the practice of early ambulation and discharge. It should also be noted that some of these studies presented data on morbidity other than those summarized in table 14. These morbidity data were discussed in relation to each study and have been omitted for simplicity. They do not alter the general conclusions of this discussion.

The authors of all of these studies concluded that the experimental short LOS could be employed with safety, because there was no statistically significant increase in morbidity or mortality. As previously discussed, only in the case of psychiatry does this conclusion seem justified by the data. For the other areas, the most distressing problem is the lack of statistical power to detect clinically significant increases in morbidity or mortality. None of these studies had less than a 25-percent chance of making a Type II error in accepting the null hypothesis of no difference if in fact a clinically significant difference existed. Thus, the statistical power in these studies was always less than 75 percent. Increasing sample sizes would increase the power (and decrease the chance of a Type 11 error), but in clinical trials, sample size is often kept small in case of harmful effects to patients.

The problem in interpreting negative clinical trials has been reviewed by Freiman and colleagues **(52).** They were concerned about the possibility that a new treatment might be abandoned

Table 14.—Summary of RCTs

	MI	Elective surgery	Low-risk obstetrics	Low birth weight infants	psychiatry
1. Number of methodologically sound RCTs	3	5	1	2	6
2. Number of studies excluding the elderly.	2	4	NA[a]	NA[a]	0
3. Average percent of screened patients accepted into study	55	68	24	100	68
4. Mortality: number of studies where					
a. E > L[b]	0	nd[c]	nd	0	nd
b. L > E	3	nd	nd	0	nd
c. L = E	0	nd	nd	2	nd
5. Power: number of studies with power > 0.75 to detect 50°/0 increase in mortality	0	0	0	0	nd
6. Morbidity : number of studies where					
a.. E > L	1	4	0	0	0
b. L > E	1	1	1	1	5
c. L = E	1	0	0	1	1
7. Power: number of studies with power > 0.75 to detect 50°/0 increase in morbidity	0	0	0	0	nd

[a]NA . not applicable
[b]E . early discharge group, L = late discharge
[c]nd = data not reported or cannot be derived

after a negative clinical trial in which no difference was observed between study and control groups. In that circumstance, a Type II error might lead to a failure to appreciate a beneficial effect of a new treatment simply because the sample size was too small to demonstrate a statistically significant difference.

The problem in the present analysis is just the reverse—a concern with missing a harmful effect of early discharge. A negative RCT in this situation might mistakenly conclude that early discharge was safe, when in fact, the study could not detect a clinically significant harmful effect due to small sample size. All of the RCTs except the psychiatric studies have some, degree of this problem in evaluating differences m mortality. In elective herniorrhaphy, low-risk obstetrics, and even low birth weight infants, the rates of reported mortality are so low that very large sample sizes would be required in order to have any reasonable chance of observing even large differences among different treatments, The sample sizes required for MI studies are somewhat less, but still greater than those employed in the reported RCTs. The same arguments apply to the morbidity measures reported.

In conclusion then, one cannot exclude the possibility that patients with uncomplicated MIs, elective surgery, uncomplicated deliveries, or low birth weight infants may experience a clinically significant increased risk of mortality or morbidity when discharged earlier than more conventional treatment practices. The RCTs reviewed do establish persuasively the lack of extremely large negative or positive effects on health outcomes of early discharge in these clinical areas. Studies with larger sample sizes will be required to evaluate the possibility of small to moderate effects.

In addition to failing to establish unequivocally the safety of early discharge, the medical literature also fails to shed additional light on the meaning of regional LOS variations. It cannot be inferred that western lengths of stay are as safe as eastern, because a clinically significant adverse effect of early discharge cannot be ruled out in those clinical areas studied. Further, the precise ways in which eastern and western physician practices differ are unknown. Since these practices have not been explicitly compared in any of the reviewed RCTs, it is also unknown whether the results of the RCTs would differ between regions. How long are MI patients kept at bed rest in the East and the West? What kind of anesthesia is used in herniorraphy patients in the East and the West, and when are they first ambulated? When are low birth weight infants sent home in the East and the West and are the discharge criteria (implicit or explicit) at all similar to those used in the RCTs summarized here?

Moreover, only a limited number of clinical areas have been studied. The LOS differences between the East and the West are pervasive-found in virtually every diagnostic group. As demonstrated, the medical issues involved in the relative safety of early discharge vary enormously from one clinical condition to another: from whether a 4½ pound neonate is feeding adequately to whether a 60-year-old man with an uncomplicated MI should be allowed out of bed on the sixth or seventh day of his hospital stay. This heterogeneity precludes generalizing the results of a few RCTs in a few clinical areas to other patients with other conditions. No information exists at all that would allow conclusions to be drawn concerning the relative safety of various lengths of stay in these unstudied clinical areas, which comprise the vast majority of hospital patients.

A recent review by Berk and Chalmers (15) evaluated data in the literature for evidence that outpatient care could be safely and economically substituted for inpatient care. They reviewed some of the same RCTs on early discharge that were reviewed in this case study. They concluded that many of the studies were methodologically flawed and that little support was available for the proposition they set out to investigate. Although the analysis in this case study did not address the issue of savings attributable to short LOS, a similar conclusion has been reached with respect to the safety of short lengths of stay.

IMPLICATIONS FOR RESEARCH

Two overall suggestions for future research emerge from this study. The potential economic significance of regional LOS variations combined with a general lack of understanding of their health implications make a large-scale study of current medical practice very important. A diverse sample of clinical conditions should be selected for which large differences in regional LOS exist and for which a significant proportion of the inpatient population is affected. A protocol should be designed to sample patients in several eastern and western localities . Thus, severity of illness can be precisely measured, regional differences in physician practices can be recorded within severity classes, and regional differences in outcomes can be assessed. Only a study of this nature can remove the mystery concerning the meaning of regional LOS variations.

The second research suggestion concerns future RCTs designed to test the efficacy of early discharge. The central problem is that very large RCTs will be required to address definitively many of the remaining questions concerning the safety of early discharge, At the same time, some of the risks that such studies would be designed to assess are quite small. The more infrequent the event, the larger the sample size, and the more expensive the study needed to detect changes in its incidence. A study is needed to set research priorities in this area. In which clinical areas are additional data on the safety of early discharge most critically needed? This question should be evaluated both from the perspective of the magnitude of the risk involved and the potential size of the economic benefit to be obtained from shortening LOS. Such a study should survey all clinical conditions for which substantial regional differences in LOS exist and should not be limited to a consideration of those areas in which studies have already been done.

IMPLICATIONS FOR POLICY

What are the policy implications of this review? Should western lengths of stay serve as standards and somehow be enforced on the rest of the country? As noted in this case study, with the exception of psychiatric hospitalization, medical research has thus far failed to exclude the possibility that early discharge is harmful. Only a handful of clinical conditions have been studied, frequently excluding substantial segments of the population. The most common finding of these studies is that for the outcome measures employed, there *is* no statistical difference between early and late discharge groups. Faced with this lack of definitive data, what conclusions may be drawn? The answer depends somewhat on where one considers the burden of proof to lie. Must proponents of early discharge prove that it is beneficial, or at least harmless? Or must proponents of longer hospital stays prove that early discharge is harmful? Physicians are likely to adopt the first position, while policy makers may prefer the second.

This question is made all the more difficult to answer, because the economic benefit of early discharge programs is difficult to specify. If the economic benefit of shortening LOS were indisputably large, then the case for early discharge would be much stronger. As already discussed, however, it is far from clear how reducing LOS produces monetary savings, either to society or to government health care programs. At one extreme, LOS may be shortened by uniformly eliminating days at the end of hospital stays through early discharges. If, however, these patients are replaced by patients who require more services per day, the net effect of such a program of LOS reduction would be to increase total costs.

On the other hand, if LOS is reduced to the point where an individual hospital experiences a large decrease in its occupancy rate and if this deficit is not replaced, the hospital may close entirely or convert some portion of its beds to another,

less costly use. In this circumstance, reducing LOS could result in a net reduction in total costs. Thus, whether LOS reductions actually save money will depend heavily on precisely how they are brought about.

How to reduce LOS so that such savings occur is an area totally unapproached by current health services research. No U.S. study has even attempted to address this question. The one British study that did address the issue (3,4) looked only at the effect of early discharge on the costs of care for the patients in the study. No attempt was made to assess the effect of early discharge on total health care costs to the community. Therefore, in the face of limited data, uncertain benefits, and possible harm, the case for early discharge seems unconvincing at best.

Does this conclusion change if one considers the perspective of the Medicare and Medicaid programs? Many of the most rigorously designed studies in the medical literature that tested the effects of early discharge programs excluded the elderly from their study populations. No study has been performed that specifically and rigorously examined the effect of early discharge on a Medicare or Medicaid population, These populations are sufficiently different from the general populations ordinarily represented in the medical literature that one must be hesitant before extrapolating results from the one to the other. Thus, the data available to judge the effects on health outcomes of early discharge for the poor and the elderly are even more scarce than for the rest of society. It is therefore even more difficult to make a strong argument in favor of early discharge in the context of the Federal health programs.

Hospital LOS varies greatly from place to place across the United States. This review has demonstrated that very little is known concerning the health consequences of these variations. Only future research along the lines described earlier in this chapter can provide the basis for rational judgments about whether hospital stays are too long in the East, too short in the West, both, or neither. Existing data cannot exclude any of these possible interpretations.