Returns to Scale in the Production of Hospital Services

by Ralph E. Berry, Jr.

The primary purpose of this article is to investigate whether or not economies of scale exist in the production of hospital services. In previous studies the results have implied the existence of economies of scale, but the question has not been satisfactorily resolved. The factor most responsible for clouding the issue is the overwhelming prevalence of product differences in the outputs of hospitals. In this study a method which avoids the problem of product differentiation is developed.

The analysis strongly supports the conclusion that hospital services are produced subject to economies of scale.

Introduction

For a number of years the costs of hospital services have displayed a marked tendency to increase both absolutely and relative to other goods and services. This phenomenon has stimulated a considerable amount of public concern and professional analysis.

The relationship between the cost of producing hospital services and the level of output of the individual hospital is central to the complex nature of hospital costs. This relationship, which is often somewhat inaccurately referred to as the "cost/size relationship," has two dimensions, each of which has important implications in its own context.

In the short-run, inasmuch as the size of the individual hospital is relatively fixed, and much of the hospital's budget is in the nature of fixed costs, the average cost of services produced will vary as these fixed costs are spread over fewer or more patients. If fixed costs are a significant proportion of the total costs of operation then the average cost will fall substantially as the level of output increases. Empty hospital beds are expensive. A number of analysts have addressed themselves to this aspect of the relationship between the cost of producing services and the level of output of the hospital [1-4]. Feldstein, for example, has estimated that the marginal, or variable, cost of a patient day was from 21% to 27% of the average cost, depending on the type of services involved [4]. There is evidence that on the average an empty hospital bed is about 75% as expensive as an occupied bed [5, 6], which implies that marginal cost is approximately 25% of average cost.

The relationship between short-run average costs and the level of operation of the hospital has important implications for policies concerned with the level of utilization, construction of excess capacity, and duplication of facilities within a hospital area. These are important problem areas, and much remains to be done before a claim can be made that the community is getting the most out of its existing stock of hospital facilities. On the other hand, the nature of short-run hospital costs is better understood than the nature of long-run hospital costs, which is the second dimension of the relationship between the cost of producing hospital services and the level of output of the individual hospital.

If new facilities are to be built, or existing facilities expanded, what is the appropriate size of a hospital? Is there an optimal size for hospitals? Answers to these questions require a knowledge of the nature of the internal cost structure of hospitals in the long-run. Are there economies or diseconomies of scale in the production of hospital services?

Returns to Scale

Conceptually, the question of returns to scale is straightforward. In the production of hospital services, just as in the production of any good or service, certain resources, or factors of production, are employed as inputs in order to obtain a given output. Returns to scale have to do with the relationship that exists between the inputs and the output. Specifically, the question of returns to scale has to do with what happens to the level of output as the number of inputs increases. If, when the inputs are increased in equal increments output grows at a constant rate, there are constant returns to scale (constant costs); if output grows at an increasing rate, there are increasing returns to scale (decreasing costs); and if output increases at a decreasing rate, there are decreasing returns to scale (increasing costs). What happens, for example, to the level of output when all inputs are doubled? Does output double? Does output increase by more than twofold? Or does output increase but less than double?

When output increases more than proportionately to inputs there are economies of scale. When output increases less than proportionately to inputs there are diseconomies of scale. When output increases in the same proportion as inputs there are neither economies nor diseconomies of scale.

Economies of scale can be a consequence of the specialization of factors of production. Division of labor, for example, may permit greater specialization resulting in increased productivity. It seems reasonable to except that division and specialization of nursing services in hospitals would result in economies of scale in this sense. Economies of scale may also arise from the use of certain indivisible factors of production. To the extent that certain "lumpy" inputs were required in the production of hospital services, economies of scale would result from the combination of factors in more efficient proportions concomitant with higher levels of output. A number of facilities might be suggested which seem to be indivisible in this sense. It is probably impossible, for example, to construct "one-half" of a pathology laboratory.

Diseconomies of scale are usually associated with the diseconomies of exceedingly large-scale management. If a hospital becomes large enough, for example, the burden of administration may become disproportionately great.

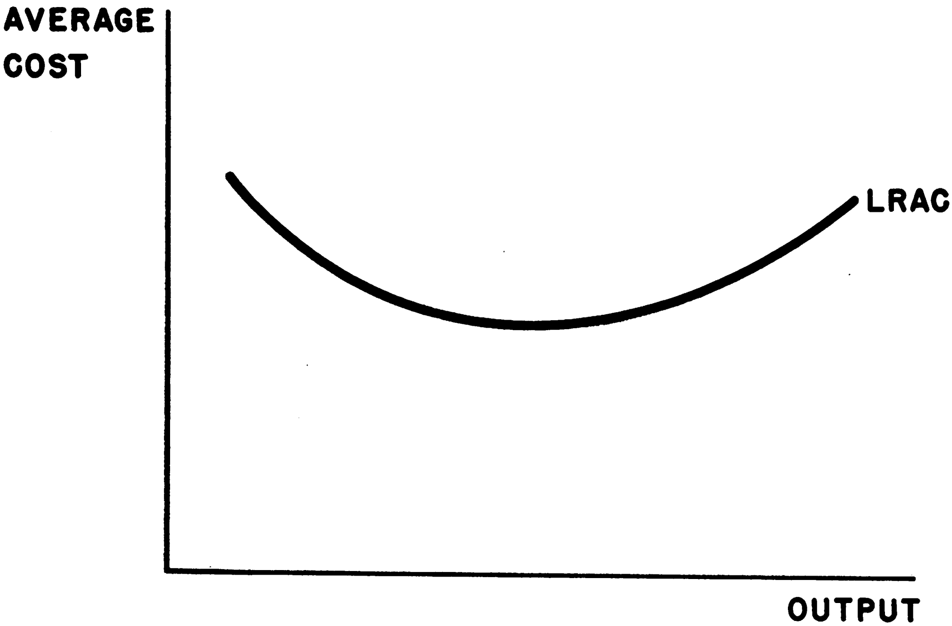

Fig. 1. A representative long-run average cost curve.

Traditionally, long-run average cost curves have been depicted as U-shaped (Figure 1). The rationalization of the U-shape is that at lower levels of output the economies of scale dominate but that eventually all the economies of increasing scale will have been fully exploited and the diseconomies of exceedingly large scale tend to dominate. The concept of an optimal size follows from the balancing of these forces. If average cost first falls and then rises as the level of output increases, the optimal size is that which coincides with a minimum average cost.

The question of whether or not economies of scale prevail in the production of hospital services is significant from the point of view of practical policy considerations for related reasons. On the one hand, the prevalence of nonprofit enterprises in this industry implies that certain incentives for efficiency which are present in industries characterized by the profit motive are missing or inoperative in the hospital industry. On the other hand, forces external to the market influence the construction of new facilities and the expansion of existing facilities. The federal Hill-Burton program, state and local agencies, and area planning groups are in a position to affect the size of hospitals. Rational planning requires some indication of the relationship between hospital size and hospital cost.

A superficial analysis of hospital costs could lead to the conclusion that hospital services were produced subject to diseconomies of scale—large hospitals do have higher average costs than small hospitals. This conclusion is misleading because it rests upon an erroneous implicit assumption of product homogeneity. Whatever else may be characteristic of hospitals, they do not produce homogeneous products.

Cost Analysis and Product Differentiation

Unfortunately, cost analysis is more straightforward in theory than it is in practice. Because what is desired is a determination of the relationship between costs of production and the level of output, it would seem that empirical analysis would require simply a number of observations of different levels of output and their respective costs. Traditionally, two empirical approaches have been utilized to determine what happens to the cost of output as the level of output increases. One approach has been to select a particular firm and study its costs and output over a period of time. This is the time-series approach. The other approach, the more common of the two, is to select a given time period and observe the relationship between cost and output for a large number of firms. This is the cross-section approach. Both methods of analysis are subject to certain advantages and disadvantages[7]. In either case, cost analysis is often complicated by variations in product quality and by inability to segregate costs by product for multi-product firms. Both of these complications are particularly pronounced in any analysis of hospital costs.

Both the quality of hospital services and the varying proportions of specific services produced deserve special consideration in the context of cost analysis.

Variations in the quality of a particular hospital service undoubtedly exist among hospitals and they probably exist within a single hospital as well. Two different hospitals may provide the service of an appendectomy, for example. Although the appendectomy is the unit of output produced, the quality of the product may vary among the hospitals that produce it. In this sense, the quality of the product will vary with the quality of resources and the level of technology. The patient who receives treatment in a large urban teaching hospital undoubtedly purchases a "better" product than he would in a nonteaching hospital. The quality of the services of teaching hospitals is affected favorably by the availability of more highly skilled human resources, more modern facilities, and more up-to-date technology. Teaching hospitals probably therefore produce a uniformly better range of services than most nonteaching hospitals and there is probably a systematic pattern of higher quality in larger hospitals. Usually, higher quality is associated with higher costs of production and conceptually cost analysis requires standardization for quality variations. Unfortunately, there is no feasible way to quantify this particular dimension of quality for hospital services and its influence on costs cannot be estimated directly. It is necessary, therefore, to modify any empirical analysis of hospital costs to account for differences in the quality of the output.

A further complication for hospital cost analysis results from the fact that output data do not reflect differences in the scope of services provided in different hospitals. Hospitals are essentially multi-product firms and total costs are not generally segregated for the different products produced. Thus, for example, two hospitals which produce the same range of services may have different average costs because they produce different proportions of these services, or two hospitals may have different average costs because they produce different ranges of services. It is conceivable that two hospitals might have the same level of output as measured by patient days, admissions, discharges, or some other unit of output and yet have quite different total costs because they have not produced the same "product." Because the nature of available data often requires that the analysis of hospital output proceed as though hospitals produce a single product, it is necessary to treat differences in case-mix and complexity of scope of services as elements of product differentiation. As the range of diagnoses a given hospital is capable of treating is directly related to the spectrum of facilities and services available in that hospital, the availability or nonavailability of facilities and services is probably a reasonable approximation of product differentiation in the form of more or less complex scopes of services.

The American Hospital Association reports the availability and nonavailability of some 28 facilities and services[8]. It is possible, therefore, to determine whether a given hospital does or does not have such facilities as a blood bank, pathology laboratory, or X-ray equipment. Large hospitals tend to have higher average costs, but they also tend to have more facilities and services available and therefore tend to produce a more complex scope of services. To determine the relationship between hospital size and hospital cost, it is necessary to allow for the fact that cost, size, and the scope of services provided are highly interdependent.

In a number of recent studies an attempt has been made to estimate the long-run cost functions of short-term hospitals. Although their findings have been consistent more often than not with the existence of economies of scale, the question is far from resolved. Klarman has summarized earlier attempts to measure cost relationships[9].

Recent Attempts to Measure Cost Relationships

Using data for some 60 hospitals ranging in size from 48 to 453 beds, Feldstein estimated a long-run total cost curve that was consistent with economies of scale[10]. Using a linear regression model he obtained the following result: Total operating expenses = $267,692.00 + $22.86 × PD where PD is patient days per year.

Thus, Feldstein found constant, long-run marginal costs which, because of the statistically significant positive constant term in the regression equation, are below long-run average costs. This relationship between marginal and average costs means that long-run average costs must be falling over the range of hospital sizes included in his sample. Although Feldstein did not adjust the data in his sample for differences in product mix, he qualified his results as follows:

> . . . large hospitals do have higher total costs, but as size increases the increase in total costs is constant. Since no allowance has been made for the number of services offered by each hospital in this sample, there is a large upward bias in the results. This bias is caused by the increase in the number of services and not by increases in the size of hospitals. Therefore, if the number of services could be allowed for, one would expect the total costs in hospitals of increasing size to increase at a *decreasing* rate . . .[11]

He could have added that this is even more likely for increases in the complexity of services.

A positive constant in the long-run total cost function might be indicative of the presence of certain indivisible factors of production. To the extent that Feldstein's result implies that certain "lumpy" inputs are required in the production of hospital services, economies of scale resulting from the combination of factors in more efficient proportions at higher levels of output are implied.

Of course, a positive constant could be explained empirically by the lack of observations at exceedingly low output levels. In the Feldstein study, for example, the smallest observation on the level of output was in excess of 15,000 patient days. If all factors of production are variable and infinitely divisible, the long-run total cost curve must pass through the origin. The LRTC in Figure 2 is drawn to exhibit decreasing costs up to the output level B and increasing costs beyond. The economies of scale implicit in

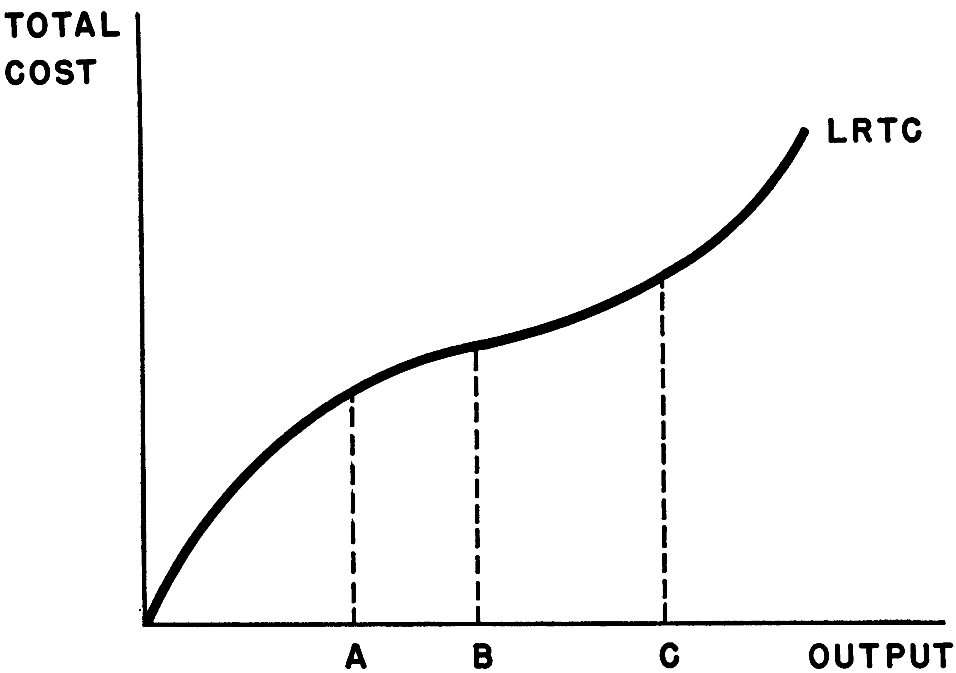

Fig. 2. A representative long-run total cost curve.

Figure 2 are those resulting from the specialization of factors. If most of the observations were from the range A to C, a linear estimate of LRTC would certainly contain a positive constant.

Whether the positive constant is attributable to the existence of indivisible factors in the production function or to a misspecification of the cost function, the implication is the same—hospital services are produced subject to economies of scale.

In a study of 72 Massachusetts community hospitals, ranging in size from 30 to 330 beds, Ingbar and Taylor derived a nonlinear estimate of the long-run average cost curve that had an inverted U-shape [12]. Their results suggest that hospital services are produced subject to diseconomies of scale up to levels of output consistent with a hospital size of approximately 200 beds and economies of scale beyond. This uniquely shaped cost curve can perhaps be rationalized by the interdependence of hospital cost, hospital size, and the scope of services provided. In effect, it seems plausible that a given set of services may be produced subject to economies of scale but that increasing the scope of services increases the average cost per unit of output. Now, if as hospitals grow in size they not only expand the scale of original services but also add new services to their spectrum, the net effect on their

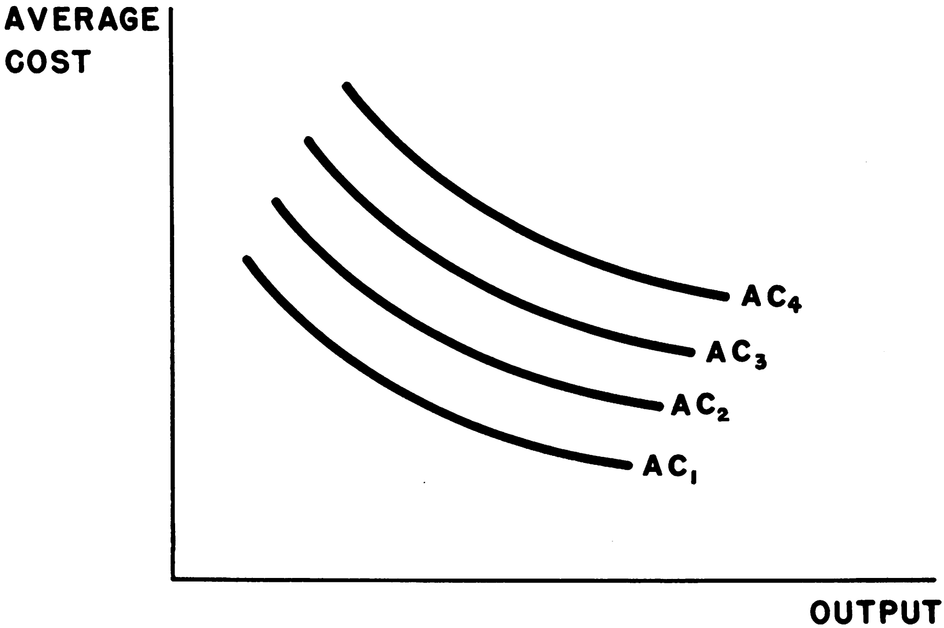

Fig. 3. Average cost curves for different scopes of services.

average cost will depend on the balancing of the negative and positive influences of scale and complexity.

The Ingbar-Taylor results may be viewed as evidence that increased complexity of the scope of services tends to be the dominant influence on average costs up to a hospital size of approximately 200 beds but that beyond that size economies of scale for specific services exert a more pronounced influence. In essence, this argument is consistent with the cost curves depicted in Figure 3. A number of cost curves are presented to represent different complexities of services produced. The average cost curves are drawn to exhibit economies of scale for any level of services but as the scope of services becomes more complex, the average cost curve is higher.

This line of analysis suggests that hospitals should be considered as small or large not only in an absolute sense but also relative to the scope of services that they are producing. A 200-bed hospital may be the optimal size for producing a narrow range of services but an exceedingly inefficient size for producing a wide range of services. The results of the Ingbar-Taylor study do suggest that there may be some upper limit to the scope of services that a hospital can provide and that once a hospital becomes a full service or near full service institution, increases in the level of output can be obtained for less than proportionate increases in cost, at least within the range of size included in their sample.

However the Ingbar-Taylor results are viewed, they do serve to accentuate the problems that arise in hospital cost analysis because of the interdependence of cost, size, and the scope of services.

In another study of hospital costs a research group collected extensive data covering a period of 14 years for 30 hospitals in six western New York counties[13]. In this study some key behavior patterns of hospital costs were found to be nonlinear, and marginal costs to vary with certain independent variables, particularly occupancy rate, scope of services, and technology, rather than to remain constant as a linear hypothesis specifies. The authors explicitly recognized the problem posed by the scope of services offered in different hospitals and divided their sample for certain analytical purposes into four groups according to the number of facilities and services available. They then used the ratio of ancillary expenses to total operating expenses as a variable to represent the scope of services in their regression analysis. This does not adequately adjust for product differentiation. Although the authors did not address themselves directly to the question of economies of scale, certain of their findings are relevant. They found marginal costs remained constant for changes in hospital size, but also that marginal costs declined with increases in occupancy rate. Given the relationship between size and occupancy rate that prevails in the short-term hospital industry, these two results, considered together, suggest increasing returns to scale.

Hospital Size, Occupancy Rate, and Economies of Scale

It has often been demonstrated empirically that larger hospitals have higher occupancy rates than smaller hospitals. This probably arises in part from the random aspects of the demand for hospital services. As the incidence of much of the illness requiring hospitalization is randomly distributed in any given population, patients should arrive at hospitals at random intervals. Day-to-day patient arrivals will fluctuate randomly about some average daily census which is the mean of the distribution. The occupancy rate will depend both on the mean of the distribution and the standard deviation about that mean. The dependence of the occupancy rate on the mean of the distribution is obvious. The average daily census divided by hospital size yields the occupancy rate. The dependence of the occupancy rate on the standard deviation of the distribution implicitly rests on the assumption that hospital size is chosen in such a way as to minimize the chance that patients will have to be turned away.

A simple example should illustrate the positive correlation between hospital size and occupancy rate. Suppose two contiguous geographical areas, A and B, are considered. Let X and Y represent the average number of persons requiring hospitalization in A and B, respectively. If each area were served by a hospital, the average daily census of A's hospital would be X, the average daily census of B's hospital would be Y. The variances of the two distributions can be represented by Var (X) and Var (Y). The standard deviation in each case would be the square root of the variance of the distribution, $\sqrt{\text{Var (X)}}$ and $\sqrt{\text{Var (Y)}}$.

Suppose that both areas were served by a single hospital. The average daily census of that hospital would be X + Y, since the mean of the sum of two distributions is the sum of the means of the distributions. The variance of the distribution would be Var (X) + Var (Y), since the variance of the sum of two distributions is the sum of the variances of the two distributions.[1] But, the standard deviation of the distribution would be $\sqrt{\text{Var (X)} + \text{Var (Y)}}$ which is necessarily less than $\sqrt{\text{Var (X)}} + \sqrt{\text{Var (Y)}}$. Therefore the two areas could be served as well by a single hospital with fewer beds as by the two separate hospitals with their combined number of beds. Alternatively, with the same number of beds the probability that patients would have to be turned away could be reduced. Since the average daily census of the single hospital would be equal to the combined average daily census of the two separate hospitals, the occupancy rate of the single hospital would be higher than the occupancy rates of either of the two separate hospitals or than their combined occupancy rate. Since empty beds have a positive cost, the relationship between hospital size and occupany rate results in an economy of scale in the same way that an increase in the dimensions of a pipeline leads to an economy of scale in the conveyance of gas or oil.[2]

This argument usually evokes an inquiry as to its generality. Does it mean, for example, that the United States would be best served by a single exceedingly large hospital? The answer is yes, other things equal, but, of course, other things are never equal. The relationship between hospital size and occupancy rate is probably a pure economy of scale, but other factors influence the internal cost structure of a hospital. Indivisible factors of production and opportunities for increased specialization of factors have been suggested as possible factors leading to economies of scale. On the other hand, the burden of administration has been suggested as a possible diseconomy of scale, which would probably become dominant long before a hospital reached the size necessary to serve a large geographical area, let alone the entire United States. Further, even if economies of scale to the individual hospital continued to dominate its internal cost structure, factors external to the hospital would serve to limit its size. Although the hospital

[1]Unless the two distributions are not independent of each other, in which case the variance of the sum of the two distributions would be Var (X) + Var (Y) + Cov (X,Y), i.e., there would be a covariance term.

[2]When the dimensions of a pipe are doubled, for example, the carrying capacity is more than doubled.

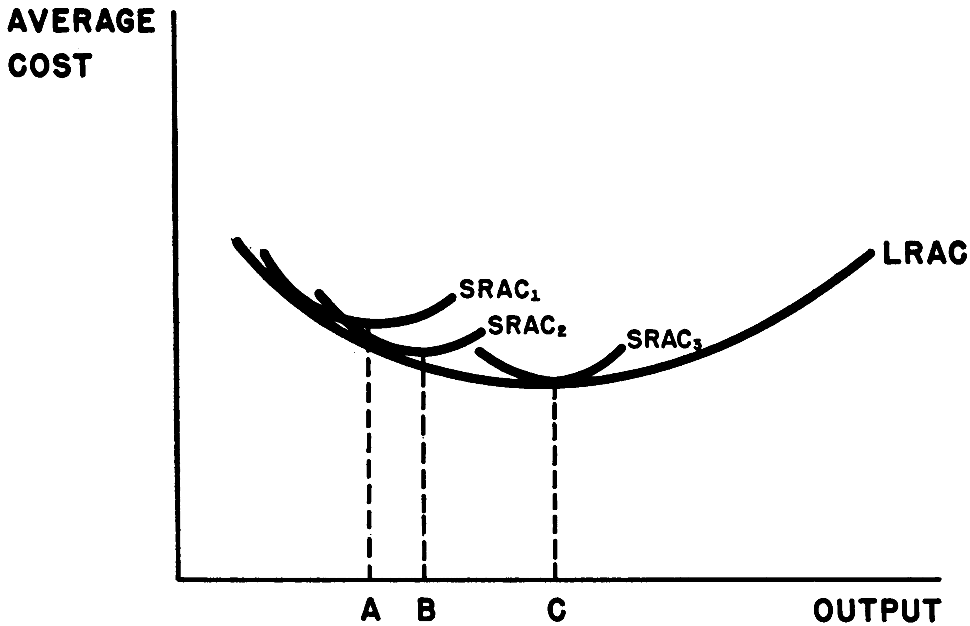

Fig. 4. The long-run average cost curve and representative short-run average cost curves.

does not have to absorb them directly, there are associated costs, including, for example, transportation, the need to treat emergency cases quickly, inconvenience, travel time for patients, physicians, and visitors, which will serve to constrain the geographical market of an individual hospital.

The relationship between occupancy rate and hospital size may be viewed further as a consequence of rational economic behavior consistent with the existence of economies of scale. If increasing returns to scale exist, the long-run average cost curve will have a downward slope over the range of the output dominated by economies of scale. Figure 4 depicts the traditional U-shaped long-run average cost curve with economies of scale prevalent up to the level of output C. Although $SRAC_1$ is constructed in such a way as to reach minimum average cost for A units of output, the optimal way to produce A is to construct a plant with minimum average cost at the level of output B, represented by $SRAC_2$, and to produce at some level below capacity, specifically at A. If minimum average costs occur at, or near, capacity, plant 3 will produce at, or near, 100 percent of capacity and plant 2 will produce at, or near, A/B percent of capacity. In general, smaller plants will operate at lower rates of capacity than larger plants. Inasmuch as bed size is an approximation of hospital capacity, and the average daily census represents the level of output, smaller hospitals should operate at lower occupancy rates than larger hospitals if hospital services are produced subject to economies of scale.

A Method of Cost Analysis Which Avoids the Problem of Product Differentiation

Although the various studies of hospital costs have more often than not implied economies of scale and certain characteristics of the hospital industry, such as the relationship between occupancy rate and hospital size, are consistent with the existence of economies of scale, the question has not been satisfactorily resolved. The factor most responsible for clouding the issue is the overwhelming prevalence of product differentiation, especially that associated with the complexity of the scope of services. If it were possible to eliminate, or to substantially reduce, the effect of product variation on hospital cost, the problem of measuring the relationship between the cost of production and the level of output would be more straightforward.

Any statistical study of hospital costs is necessarily subject to the complication of product differentiation. Time-series studies of hospital costs are frustrated by the fact that if a hospital grows in size over time, the scope of services offered by that hospital usually increases. This product variation is added to the "normal" problems of adjusting for changes in technology, factor prices, and other factors which do not remain constant over time. Cross-section analysis is also complicated by the fact that product complexity differs with hospital size. Unless one can hold the product constant, or reasonably constant, one is unable to measure the conceptually simple relationship between the level of output and the cost of producing that output.

As one of the primary purposes of this study is to determine the existence or nonexistence of economies of scale in the short-term general and other special hospital industry, it is necessary to develop an estimate, as free as possible from the influence of product differentiation, of the average cost function of producing hospital services. The optimal solution would be to find a number of hospitals which produce a perfectly homogeneous product but which vary in size. A cost curve could be estimated statistically for this group of hospitals and the question of returns to scale resolved for this particular group, at least. If it were possible to group all, or most, short-term hospitals in such a way, the question of returns to scale in the industry could be resolved. Such an optimal grouping could be obtained only from an ideal index of output, the development of which would require a type of medical audit of all hospitals that would not be feasible given present financial and/or personnel constraints. It should be possible, however, to approach the optimal solution by successive empirical approximation. The following methodology seems plausible given the practical constraints.

The total output of hospitals is usually measured by patient days. It can also be measured by admissions, discharges, or by the average daily census. Whichever measure is used, the lack of homogeneity in output data is a consequence of the varying proportion of patients with different diagnoses treated in different hospitals. A group of hospitals would have reasonably homogeneous outputs if they treated the same proportion of patients with the several diagnoses. In the absence of data concerning the exact proportion of diagnoses treated, an approximation of the scope of services provided by given hospitals is needed. The scope of services provided by a given hospital is directly related to the facilities and services available in that hospital. Although the availability of a particular facility or service does not establish the extent of its use, nonavailability certainly establishes its nonuse. For example, the availability of an operating room does not indicate what proportion of total patient days are surgical patient days, but the nonavailability of an operating room does suggest that total patient days include no surgical patient days. Although it is impossible to separate those hospitals in which 50% of patient days include surgical services from those hospitals in which 40% of patient days include surgical services, it is possible to separate hospitals with some from those with none.[3] This is certainly a step in the right direction. For any given facility or service, two groups of hospitals could be formed—those having the facility or service and those not having it. This process could be repeated for all 28 facilities and services reported by the American Hospital Association. Finally, all hospitals with exactly the same facilities and services would form a group.

This method of grouping hospitals with identical facilities and services is a second-best approximation for grouping by product homogeneity. In essence, a group of hospitals with identical facilities would be assumed to be producing a relatively homogeneous product. If groups of short-term hospitals that have exactly the same facilities and services are isolated, the relationship between average cost per unit of output and the level of output within these groups can be measured directly. Such an analysis should shed light on the question of whether or not economies of scale exist in the production of hospital services.

The primary source of data for this analysis was the American Hospital Association. Sufficient data were available to include 5,293 of the 5,684 nonfederal, short-term general and other special hospitals registered in 1963. First, groups of hospitals with identical facilities as listed in Table 1 were obtained by sorting according to the availability or nonavailability of the 28 facilities and services. The skewed distribution of the data is to be expected. As 28 facilities and services are involved, the probability of any one hospital having a particular combination of facilities and services is exceedingly small. Given the probability that any one hospital will have a specific combination of facilities and services, the likelihood that there will be large groups (many

[3]This ignores the possibility that a hospital may have but not use an operating room.

Table 1. NONFEDERAL SHORT-TERM GENERAL AND OTHER SPECIAL HOSPITALS GROUPED BY IDENTICAL FACILITIES AND SERVICES FOR 1963

Group Size	Number of Groups	Number of Hospitals
1	2812	2812
2	323	646
3	114	342
4	43	172
5	32	160
6	19	114
7	14	98
8	12	96
9	10	90
10	4	40
11	6	66
12	7	84
13	2	26
14	1	14
15	5	75
16	1	16
17	1	17
18	2	36
19	2	38
20	1	20
21	2	42
—		
24	1	24
—		
26	1	26
—		
33	1	33
—		
45	1	45
—		
69	1	69
—		
92	1	92
		5293

hospitals sharing a particular combination) is such as to suggest that skewness that was actually observed. Fortunately, however, it was possible to obtain a significant number of groups of reasonable size. Each group is made up of a number of hospitals which produce hospital services with the same facilities and services. For example, the largest single group contains 92 hospitals which have seven identical facilities and services. The output of these 92 hospitals is assumed to be reasonably homogeneous, and product variation should exert a minimum influence on the relationship between the cost and the level of output of these hospitals. The same is true for the group of 69 hospitals, the group of 45 hospitals, and others. It was, therefore, possible to analyze the cost/output relationship of a significant proportion

Table 2. SUMMARY OF STATISTICAL ANALYSIS OF THE RELATIONSHIP BETWEEN AVERAGE COST AND OUTPUT IN SHORT-TERM GENERAL AND OTHER SPECIAL HOSPITALS

Group	Group Size	No. of Facilities and Services in Common	Correlation Between Average Cost and Output	Regression Equation Average Cost =	T Statistic For Beta	Degrees of Freedom
1	92	7	−.14	29.41 − .0003 PD[1]	1.30	90
2	69	6	−.37	35.72 − .0008 PD	3.27	67
3	45	7	−.09	29.05 − .0003 PD	.56	43
4	33	8	−.23	35.05 − .0006 PD	1.31	31
5	26	8	−.21	31.83 − .0004 PD	1.05	24
6	24	8	−.21	37.02 − .0006 PD	.98	22
7	21	8	−.24	35.62 − .0007 PD	1.08	19
8	21	8	−.35	35.21 − .0007 PD	1.63	19
9	20	5	−.02	27.48 − .0001 PD	.10	18
10	19	7	−.38	42.41 − .0015 PD	1.68	17
11	19	7	−.60	36.71 − .0012 PD	3.09	17
12	18	9	−.28	38.95 − .0005 PD	1.17	16
13	18	8	−.69	35.92 − .0004 PD	3.81	16
14	17	9	−.41	33.43 − .0006 PD	1.75	15
15	16	7	−.37	27.83 − .0004 PD	1.51	14
16	15	26	−.59	58.63 − .0001 PD	2.63	13
17	15	8	−.24	33.50 − .0005 PD	0.88	13
18	15	8	−.15	33.58 − .0003 PD	0.55	13
19	15	8	−.31	31.48 − .0008 PD	1.17	13
20	15	8	−.36	37.86 − .0008 PD	1.41	13
21	14	7	+.18	23.93 + .0003 PD	0.62	12
22	13	13	−.38	44.11 − .0003 PD	1.37	11
23	13	7	+.21	22.96 + .0003 PD	0.73	11
24	12	11	−.80	44.35 − .0010 PD	4.28	10
25	12	11	−.48	37.46 − .0006 PD	1.73	10
26	12	10	−.42	33.30 − .0003 PD	1.46	10
27	12	9	−.26	30.06 − .0004 PD	0.84	10
28	12	8	−.17	30.99 − .0003 PD	0.56	10
29	12	8	−.30	38.14 − .0005 PD	1.01	10
30	12	5	−.47	43.96 − .0020 PD	1.69	10
31	11	9	−.41	37.48 − .0005 PD	1.35	9
32	11	9	−.33	34.80 − .0004 PD	1.06	9
33	11	9	−.17	31.02 − .0006 PD	0.52	9
34	11	9	−.36	32.05 − .0003 PD	1.15	9
35	11	9	−.34	39.45 − .0013 PD	1.10	9
36	11	12	−.16	44.23 − .0001 PD	0.49	9
37	10	10	−.66	49.17 − .0016 PD	2.46	8
38	10	9	−.30	32.34 − .0004 PD	0.89	8
39	10	8	+.82	18.46 + .0016 PD	4.04	8
40	10	22	+.18	34.90 + .0001 PD	0.52	8

[1]Total patient days

of the short-term hospitals within groups that produce relatively homogeneous products. After the problem of product differentiation was overcome, it was possible to deal with the question of returns to scale directly. Table 2 summarizes the results of the analysis.

These data overwhelmingly support the conclusion that hospital services in the short-term hospital industry are produced subject to economies of scale. The relationship between average cost and the level of output was analyzed statistically for 40 groups of hospitals by estimating a linear regression equation of the form $y = a + bx$ with average cost as the dependent variable and total patient days as the independent variable. The analysis included all groups with ten or more hospitals which represent almost 15% of all short-term hospitals. The correlation coefficient between average cost and total patient days is negative in no less than 36 of the 40 groups. The functional relationship between average cost and the level of output, represented by the regression equations as listed in Table 2, reflects the same general result.

The average cost curve of hospitals producing a given scope of services, as measured by the availability of facilities and services, declines as output increases in 36 of the 40 cases analyzed. Furthermore, the negative relationship is found for all groups which include 15 or more hospitals, and the only positive statistically significant relationship occurs in a group of ten hospitals.

Not all of the negative relationships are statistically significant, however. One measure of significance often employed in regression analysis is the nonparametric test in which the beta coefficient is compared with its standard error. If the beta coefficient is greater than its standard error (if, in other words, the t-statistic for beta is greater than 1.0), the variable contributes to the corrected coefficient of determination (R^2) and is significant in helping to explain the functional relationship. For 26 of the 36 equations which display a negative relationship between average cost and the level of output, the beta coefficients are greater than their standard errors. More significant, however, is the overwhelming number of negative relationships. There is very small likelihood that such a large absolute and relative number of declining average cost curves would be found if economies of scale did not in fact prevail.

Conclusion

This analysis of hospitals producing a relatively homogeneous product supports the conclusion that services are produced subject to economies of scale in the short-term general and other special hospital industry. Cost, hospital size, and the scope of services provided are highly interdependent, but for a given scope of services the average cost of production tends to be lower for large hospitals than for small hospitals. The federal Hill-Burton program, state and local agencies, and area planning groups must consider the existence of economies of scale in the production of hospital services when planning the supply of hospital facilities.

The question of returns to scale and the interdependence among hospital size, hospital cost, and the complexity of the scope of services offered assume more significance under the recent social security amendments which provide for medical care for the aged. The legislation authorizes the Social Security Administration to reimburse hospitals which treat elderly patients "according to cost." The federal government thus joins Blue Cross, private insurance companies, and others in a group of third party payers who reimburse hospitals for the services received by individual patients. The actual out-of-pocket expenditures by most patients are but a fraction of the total bill; in the case of many it is zero[14]. This phenomenon in the market for hospital services means that the traditional market incentives which promote efficiency are missing. The policy of reimbursing hospitals according to costs reinforces the unqualified desire for "Cadillac only" medical care. Unless a system of reimbursement is devised which includes certain institutional incentives that can serve to replace the missing market signals the result may be excessive duplication of activities and consequent inefficiency in an ever-increasing spiral. Conscious attempts should be made by policy makers to develop a system of reimbursement that will stimulate expansion designed to take advantage of economies of scale, but will not result in expansion to increase the scope of services unnecessarily to the extent that duplication, inefficiency, and excessive cost result.

References

1. Long, M. Efficient use of hospitals. *The Economics of Health and Medical Care* (p.211). Ann Arbor: University of Michigan, 1964.
2. Reder, M. Some problems in the economics of hospitals. *Amer. Econ. Rev. 55*:472, May 1965.
3. Weisbrod, B. Some problems of pricing and resource allocation in a nonprofit industry —the hospitals. *Jour. of Bus. 38*:18, January 1965.
4. Feldstein, P. *An empirical investigation of the marginal cost of hospital services.* Chicago: Graduate Program in Hospital Administration, University of Chicago, 1961.
5. Brown, R. The nature of hospital costs. *Hospitals, J.A.H.A. 30*:36, April 1, 1956.
6. Brown, R. What do we mean by hospital costs, *Hospital Forum (So. Cal.) 4,* May 1961.
7. Johnston, J. *Statistical cost analysis.* New York: McGraw-Hill, 1960.
8. American Hospital Association, *Hospitals, J.A.H.A. 38*:15 Part 2 (guide issue), 1964.
9. Klarman, H. E. *The economics of health* (p.106). New York: Columbia University Press, 1965.
10. Feldstein, *op cit.,* p.62.
11. Feldstein, *op. cit.,* p.64.
12. Ingbar, M. L. and L. D. Taylor. *Hospital costs: a case study of Massachusetts.* Harvard University Press, 1967.
13. Fitzpatrick, T. B., S. R. Gottlieb and G. C. Wirick. *The nature of hospital costs.* Ann Arbor: Bureau of Hospital Administration, University of Michigan, 1964.
14. Rosenthal, G. The public pays the bill. *Atlantic Monthly 218*(1):107, 1966.