The Determinants of Nursing Home Operating Costs in New York State

A. James Lee & Howard Birnbaum

This paper investigates the determinants of nursing home operating costs in the State of New York during 1975. The analysis indicates that "scale" and occupancy are minimally important in determining operating cost variation. In contrast with other studies, patient- and service-mix differences are found to be important, although reimbursement grouping variables (e.g., profit/nonprofit) remain the most significant and important variables. Having estimated one of the most general cost functions to date, the authors ask what accounts for these differences if not differences in the patient mix, service mix, or input prices. If the differences are merely due to unmeasured differences in the amenity level of care or to managerial inefficiency, then it is not clear that public programs should continue to discriminate between facilities in these categories and pay the differences.

INTRODUCTION

The Medicare and Medicaid programs began paying for nursing home care in 1967. Between 1967 and 1981, the total cost of nursing home care in this country rose from $2.8 billion to $24.2 billion. In response to such rising costs, scandals, and concern about quality of care, the federal and state governments have enacted a variety of nursing home regulations. Most recently, Public Law 92-603 mandated that state Medicaid plans provide for reimbursement of skilled and intermediate care facilities on a "reasonable cost-related basis." Development of an appropriate reimbursement scheme thus depends on an understanding of the many factors

---

The analyses described in this paper were conducted as part of a much larger study of nursing home cost and reimbursement conducted for the Health Care Financing Administration (HCFA) under Contract No. 60-77-0068. The views expressed here are those of the authors and no HCFA endorsement is intended or implied.

Address communications and requests for reprints to A. James Lee, Senior Economist, Arthur D. Little, Inc., Acorn Park, Cambridge MA 02140. Howard Birnbaum is Director of Public Programs and Long-Term Care at Abt Associates Inc., Cambridge MA.

which affect nursing home costs, including input prices, output characteristics, patient characteristics, and the extent of regulations.

Our paper investigates the determinants of nursing home operating costs in the state of New York for 1975.[1] Data came from the Medicaid cost reports submitted by long-term care facilities in that year. New York was selected primarily because the New York cost reports require more detailed reporting of services and patient characteristics than is available from the cost reports in other states.

Long-term care facilities in New York have been reimbursed on a prospective basis since 1971. Roughly speaking, in 1975, the rate was set in advance as the lower of the actual inflation-adjusted per diem cost in an earlier year and a cost ceiling computed as the 60th percentile weighted average for the comparison group. Facilities were grouped by type (SNF vs HRF[2]), sponsor (nonprofit vs. profit vs. public), bed size (0–49, 50–99, 100–199, 200–299, and 300+), and region (West, Rochester, Center, Northwest, North Metro, Long Island, and New York City), except that groups with only few cases were combined with others.

Conceptual Perspective

The conventional derivation of a cost function has cost as a function of the output level or quantity (Q) and the relative prices of inputs (P_1, $P_2, \ldots, P_n$), as shown in Equation (1).

$$C = f(Q; P_1, P_2, \ldots, P_n) \tag{1}$$

The cost function is considerably more complicated in the case of the nursing home industry. The "output" or "product" of nursing homes cannot be meaningfully measured as a single quantity. While patient days of care is a useful indicator of the output level, it is only one of many dimensions describing relevant differences between outputs of different facilities. In particular, the number of patient days per se reflects neither variation in the quality of care nor variation in the characteristics of patients receiving care—it is likely to be more expensive, other things equal, to care for a nonambulatory patient than for an ambulatory patient. Thus, the level of output Q in the nursing home cost function is properly a vector of characteristics (Q_1, $Q_2, \ldots, Q_n$), including the number of patient days but also including patient, service, and quality characteristics.

In addition, the nursing home industry is a regulated industry. If it were unregulated there would be no reason to distinguish facility type

(SNF vs. HRF), since relevant differences would be captured by patient and service variables. However, the state and federal governments have imposed differential staffing and other requirements on different facility types. To the extent that these regulations do not fully reflect underlying resident requirements, the differential regulatory requirements may imply additional differences in cost by facility type. Furthermore, the use of region, sponsorship, or bed-size categories in grouping facilities for rate setting means that historic differences in nursing home practice patterns (e.g., the "amenity" level) will be maintained as a consequence of regulation, independent of patient requirements, quality, or regional variation in wage levels.[3] Finally, public financing of long-term care may influence costs to the extent that Medicaid and Medicare provide either more or less generous reimbursement for the same level and quality of care. Alternatively, private patients, on average, may demand a higher amenity level of care than can be provided to public patients within the ceiling limitations for reimbursement.

Another more general point is the following. One may reasonably choose to have the cost function reflect alternative techniques for producing the same output. With respect to nursing homes, this can mean cost differences related to facility differences (e.g., age of facility), or differences in operating the facility (e.g., percent of services contracted).

In summary, we believe that the following is a more accurate conceptual representation of the nursing home cost function.

$$C = f(Q_1, Q_2, \ldots, Q_m; P_1, P_2, \ldots, P_n; R_1, R_2, \ldots, R_x; F_1, F_2, \ldots, F_y) \quad (2)$$

where: C is average per diem operating cost,

$Q_1, Q_2, \ldots, Q_m$ are quantity dimensions (e.g., patient or service characteristics),

$P_1, P_2, \ldots, P_n$ are the input prices (e.g., regional wage level),

$R_1, R_2, \ldots, R_x$ are variables reflecting the impact of regulation (e.g., staffing requirements, life safety codes, certificate of need, reimbursement mechanism, grouping categories, or extent of government financing),[4] and

$F_1, F_2, \ldots, F_y$ are variables denoting differences in technology or production process (age of facility).

This expression gives the determinants of long-run average costs. However, in actual estimation, the occupancy level, a measure of capacity utilization, should also be included to take account of short-run disequilibria.

EMPIRICAL SPECIFICATIONS FOR ESTIMATING THE AVERAGE OPERATING COST FUNCTION

This section indicates the operational specifications for estimating a nursing home average operating cost function from information available on the New York Medicaid cost reports for 1975. The dependent variable for purposes of estimating the cost function is the average or per diem operating expense—called AOC for Average Operating Cost. Capital-related expense (depreciation, interest, and rent), taxes, and profit are not reflected in this variable. These cost components were excluded for both practical and methodological reasons.[5]

We now proceed to the independent variables and comment briefly on the hypotheses associated with each. While our conceptual approach has suggested novel interpretations for several variables—e.g., that sponsorship is a regulatory-type variable in the state of New York—we group the variables here in more conventional fashion and note their correspondence to the conceptual categories in equation (2). In particular, we group the variables into the following major categories: (1) facility variables, (2) patient variables, (3) service variables, and (4) regional variables.

Facility Variables. Although average daily census might be a more meaningful indicator for the overall scale or level of output, we nevertheless use the number of beds (Beds) to proxy the scale of operations. The reason is as follows. Five bedsize groups—0–49 beds, 50–99 beds, 100–199 beds, 200–299 beds, and 300+ beds—are used as grouping categories for rate setting in New York. We wish to incorporate these bed-size categories directly into the analysis inasmuch as they are regulatory-type variables. In addition, we also wish to indicate noncategorical or continuous scale variables in order to test for economies or diseconomies of scale. Thus, the bed-size categorical variables may be picking up two effects—regulatory and scale. This will be explained in discussion of the results. In any event, we have chosen to use the number of beds instead of average daily census for the scale factor so that cost variation can be assessed in light of the New York grouping mechanism.[6]

Percentage occupancy (OCC) is included as a measure of capacity utilization and, as such, is thought to control for shorter-run or disequilibrium deviations from a longer-run curve. As is usual in economic analyses, we hypothesize that as firms operate closer to capacity, production is more efficient and costs are lower.

Sponsorship—for-profit (Profit), nonprofit (Nonprofit), and public (the omitted category)—are regarded as regulatory-type variables since the facilities were grouped by sponsorship for rate-setting purposes in 1975.

Facility certification—skilled nursing facility (the omitted category) or health-related facility (HRF)—is also used for grouping and thus serves a similar purpose. In addition, certification reflects the cost impact of the certification requirements imposed regardless of whether or not the required differences are needed for the actual resident population. For analytic purposes, combined facilities, i.e., facilities having both SNF and HRF parts, were split into their SNF or HRF components. Thus, we note two additional certification-related differences, whether or not an SNF is part of a combined facility (SNF-HRF) and whether or not an HRF is part of a combined facility (HRF-SNF). However, the interpretation of these certification variables is considerably different from that for the SNF vs. HRF distinction. The SNF-HRF and HRF-SNF variables are thought to reflect economies (or diseconomies) due to "joint production" and to control for possible arbitrage in the accounting of costs between facilities—e.g., systematic allocation of SNF costs to the HRF part.

The ratings from survey certification reflect the extent to which certification requirements have been met, and thus the coefficients obtained for these variables indicate the extent to which costs associated with certification have been borne by the home. To the extent that certification assures the quality of care, these rating variables also correspond to quality differences. The ratings and their distribution follow[7]:

UN	— Unsatisfactory	(0.2%)
NI	— Needs Improvement	(7.6%)
GF	— Good, Federal	(29.1%)
GS	— Good, State	(55.4%)
VG	— Very Good	(0.2%)
DF	— Deferred	(5.7%)
RNF	— Deferred, New Facility	(0.2%)
RO	— Deferred, New Owner	(0.4%)

The figures in parentheses give the relative frequency of each designation among observations with non-missing data—84.8 percent of the total sample size. Except for the deferred categories (DF, RNF, and RO), the designations are displayed in ascending order of "satisfactoriness." Unlike most states, the state certification requirements in New York are more stringent than the federal requirements. The circumstances under which the DF designation was given are not known. Furthermore, the less than satisfactory NI and UN ratings may be associated with altogether different deficiencies in different facilities—cleanliness in one, nursing in another, and food in a third. After exploring several alternatives and taking account of designations with only a few cases, the rating designations were aggregated into the following categories: "good" (GS, GF, or VG), "needs improvement" (NI or UN), and "deferred" (DF, RNF, or RO). Preliminary analyses found no meaningful difference in costs between facilities receiving either the GS or GF designation.

The percentages of patients in rooms having one (%1Bed), two (the omitted category), and three or more (%3Bed) beds, measure differences along another dimension of quality, namely, the degree of privacy among patients. These variables may also reflect differences in the technology or process of producing the same outputs. The percentage of operating expense contracted (%Contract) is included as a control for a potentially different production process. In theory, one would not contract for services unless they cost less than the alternative of having the facility produce them itself. If so, facilities with a larger percentage of contracted expense should have lower operating costs. Pragmatically, the situation may be very different, however, since contracting is an allegedly common mechanism for the fraudulent overcharging (via "kick-backs" from contractors) of publicly reimbursed patients. The age of the facility (Facility Age) is included to control for potential "vintage" effects in the substitution of capital and labor, and the associated cost implications, a source of exogenous difference in the production technology.

For each facility, we want a vector of relative prices for the various inputs (e.g., nurses, food, and equipment) used in production of long-term care. Although some regional proxy variables were available, as will be seen, there was little explicit information on the relevant price differences between facilities—only the average LPN wage (LPN Wage) within each facility.

Patient Variables. The age distribution of residents or patients is typically included in nursing home studies to proxy age-related differences in infirmity and the consequent need for differential treatment and resources. The usual hypothesis is that the cost of caring for an older patient is higher, on the average, because older persons tend to be more disabled and require more personal attention. Of course, to the extent that relevant differences in debility are measured more explicitly (e.g., percent nonambulatory), the age proxies may no longer be meaningful. In any event, we included the percentage distribution of patients age 64 and under (%Under65), age 65 to 74 (the omitted category), age 75 to 84 (%Age75to84), and over 84 (%Over84) to test for any uncontrolled effects of patient age. An earlier study [2] had indicated that female patients are more expensive. Thus, we also explore this prospect by including percent female (%Female) as a patient variable.

The New York cost reports for 1975 have a broad range of information on patient functional status—the numbers of patients requiring no assistance, partial assistance, and full assistance with each of the following: (1) walking, (2) eating, (3) bathing, (4) breathing, (5) dressing, (6) correspondence, (7) medication, and (8) massage.[8] For analytic purposes we have computed the percentages of patients requiring either some or full assistance with each of these (%Walk, %Eat, %Bathe, %Breathe, %Dress, %Correspond, %Medicate and %Massage).[9]

The percentage of patients dying each year (%Deaths), calculated as the number of deaths within the facility divided by the average daily census, might be a further indicator of patient debility or health status. On the other hand, it may also reflect, at least in some degree, the unfortunate consequences of lower-quality care. Although this variable is included, its interpretation remains ambiguous pending further study.

The average length of stay and the admission rate are very nearly nonlinear transformations of one another, given other variables already included in the model. We use an admission rate variable (AdmRate) computed as the total number of admissions divided by total patient days.[10] This definition yields a coefficient that is a straightforward estimate of the (average) fixed cost per admission.

The last patient variables available are the payer characteristics. We have calculated the percentage distribution of private (%Private), Medicaid (the omitted category) and Medicare (%Medicare) patients residing in each facility. As discussed above, we regard these as regulatory-type variables that indicate the extent to which Medicare and Medicaid are systematically purchasing more or less expensive care than private patients. Of course, any findings cannot indicate whether or not the cost differences correspond to "real" differences in the care received by patients or whether they simply reflect inefficient production or excessive reimbursement.

Service Variables. Service variables are generally regarded as proxying quality differences. We have information on service intensity and have computed the following variables for analysis: annual number of pharmacy prescriptions per patient (#Prescriptions), annual number of x-rays per patient (#X-Rays), and the average (weekly) proportion of patients receiving physical therapy (PT), speech therapy (Speech Therapy), occupational therapy (OCC Therapy), inhalation therapy (Inhal Therapy), radiation therapy (Radiation Therapy), and social services (Soc Service).[11] Facilities providing greater service intensity are presumed to be using more resources and thus are hypothesized to be more expensive.

However, the relationship between service intensity and quality is a tenuous one at best. Our approach cannot distinguish whether or not the services provided are appropriate, i.e., too few or too many given the patient mix of the home. The usual presumption has been that the more services, the better. Where this may be a reasonable assumption in the aggregate, it is probably not true with respect to all facilities.

The cost reports also identified those services provided to Medicaid patients that were included in determining the Medicaid rate. Although these variables overlap the service intensity variables to some degree, they serve a somewhat different function. They control for the accounting of service expense indicating whether or not the resource cost is reflected in the rate. Categorical (0–1) variables indicating whether or not the service was provided to Medicaid patients and its cost included in the rate were defined for the following: drugs (Drug); physical therapy (Phys Therapy); occupational therapy (OCC Therapy); speech pathology (Speech Path); special duty nurses (Special Duty); and physical services (Phys Serv). Although similar information was available for (1) dental services, (2) oxygen, (3) audiology, and (4) podiatry, we judged a priori that these service components were unlikely to be costly enough to warrant indicating them as separate service variables for analysis. Moreover, very few homes provide these services. We have aggregated these services to form an index (Other Services), ranging between zero and four, indicating the number of such services provided to Medicaid patients and included in the rate.

Regional Variables. Among the regional variables are categorical variables (Region1 to Region6) indicating the seven regions—West (the omitted category), Rochester, Central, Northeast, North Metropolitan, Long Island, and New York City—used in grouping facilities for reimbursement. Since explicit measures of regional wage and price differences are included in the cost models, these location dummies can be viewed as reflecting regional practice differences implicitly endorsed by Medicaid which groups facilities by region for rate-setting purposes. That is, they are essentially regulatory-type variables.

In addition to the facility-specific LPN wage, we also include the average wage (Retail Wage) for retail workers in the county of facility location, as a proxy for regional variation in wage and other price levels. The county-wide population density (PopDen) is also included for this purpose, under the assumption that prices tend to be higher in more populated areas.[12] This variable, however, may also reflect agglomeration economies associated with urban residence, e.g., access to specialized services such as contract laundries.

Parameter Specification. The model estimated is linear except for a few key variables. Much previous cost modeling, both with respect to nursing homes and to hospitals, has been concerned with economies and diseconomies of scale and with potential nonlinearities with respect to occupancy. We did not with to ignore these interests and yet we did not wish, as others have done, to impose an arbitrary parametric form on the relationship of these variables. We chose instead to use a piece-wise linear estimation technique, also called "grafting." This procedure, which has been largely ignored in the literature, is general in the sense that it is empirically based and permits the data to indicate a piece-wise linear approximation to an arbitrary but unknown function form (see [4]). This is accomplished by estimating separate linear segments over different ranges of the scale variables as shown in Figure 1(a). For example, if the independent variable in that figure were beds and the dependent variable average operating cost, the results would indicate economies of scale between 0 and X_1, no economies between X_1 and X_2, some diseconomies between X_2 and X_3, and substantial diseconomies above X_3. It is also possible to allow for discontinuities in the relationship as shown in Figure 1(b).

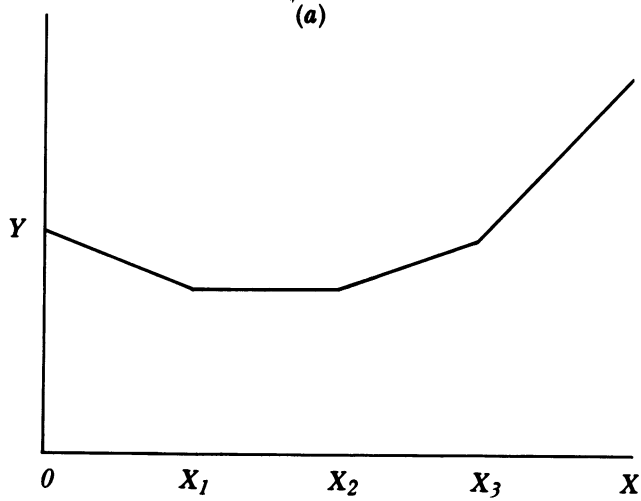
For occupancy we estimated a piece-wise relationship of the kind shown in (a). In particular, we obtained separate slope coefficients for each of the following percentage occupancy ranges: (1) 0–80% (2) 80–90% (3) 90–95% and (4) 95% and higher. The estimation involved the specification of four separate occupancy variables (OCC–OCC4).[13]

For our scale factor—the number of beds—we estimate a relationship of the type shown in Figure 1(b). Our reasoning is as follows. Since five bed-size categories (0–49, 50–99, 100–199, 200–299 and 299+) are used as grouping criteria for rate-setting purposes, we indicate four dichotomous variables (Size2–Size5) (omitting the first bed-size category) as regulatory-type variables. We then obtain a type-(b) function if piece-wise (continuous) variables are defined over the same grouping bed-size ranges (Beds1–Beds5). As already noted, scale and regulatory effects will be intertwined in the resultant relationship. We return to this issue in our discussion of the results.

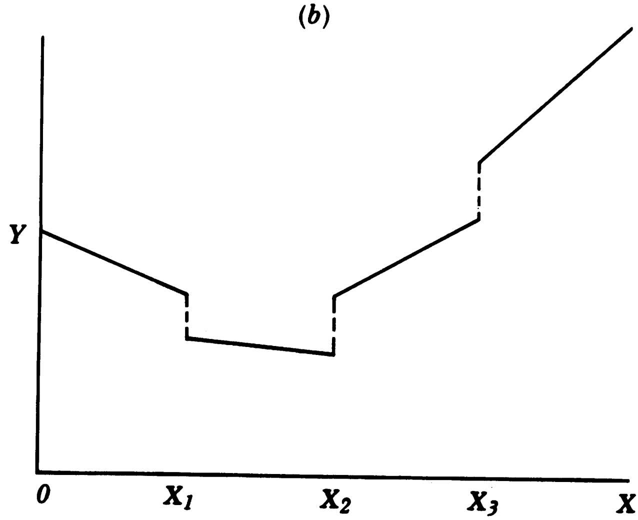
The independent variable list is lengthy and it should be apparent that the potential for multicollinearity is great. However, the reader is reminded that multicollinearity merely leads to inefficiency of parameter estimates and does not affect the "goodness-of-fit". If the purpose were only to explain cost variation using the available data, even severe multicollinearity would not be a problem. However, instability of coefficient parameter estimates may be troublesome to users of any reimbursement formula developed from such regression results. Although we have been sensitive to the potential for multicollearity in specifying the model, we see no fully satisfactory alternatives for resolving the difficulty. A sound conceptual rationale is given for including each of the variables

Figure 1: Illustration of Piece-Wise Estimation

(*a*)



(*b*)



included in the model. We sought to address the multicollinearity problem by estimating not only a complete model but also a model using stepwise selection of variables having a t-statistic of one or greater—i.e., having a coefficient estimate larger than its standard error. This corresponds to approximately a 0.30 confidence level. Although this model is not reported here, the results were surprisingly similar.

THE COST-FUNCTION RESULTS

The universe included all 671 skilled nursing facilities (SNFs) and health related facilities (HRFs) submitting Medicaid Cost Reports for 1975 in New York State. Recall that combined SNF-HRF facilities are partitioned and that such facilities are thus represented twice, both as an SNF part and as an HRF part. Hospital-based SNFs and HRFs were omitted because of a concern that such facilities might be specializing in shorter-term care and thus not be comparable to free-standing facilities. The dependent variable—average operating cost (AOC)—was formed as the sum of all payroll, dietary, laundry, housekeeping, and maintenance expenses divided by total patient days. Observations lacking any of this information were omitted. Several facilities with average operating costs less than $15 per day were also omitted. Furthermore, facilities not open for the full year were screened out. Such newly opened facilities were much less likely to be operating on their long-run cost curves. For analogous reasons we sought to eliminate facilities changing bed-size over the year. The number of usable observations, i.e., observations that remained after screening, is 504. In general, missing data were not a serious problem. However, mean values were substituted for missing values in some instances (NI, DF, %Contract, and LPN Wage).[14]

The results are reported in Table 1.[15] Beginning with the summary statistics, the coefficient of determination (R^2) is 76.8 percent, and the corrected coefficient of determination is 73.6 percent. The *F*-ratio is highly significant, and 22 of the 61 variables, or 36 percent, are significant at the 0.10 level or better. Examine now the results for individual variables and groups of variables.

Recall that both continuous (Beds1–Beds5) and categorical (Size2–Size5) bed-size variables were defined—the first explicitly to investigate scale effects and the second to control for the use of bed size as a grouping variable for rate-setting purposes (i.e., a regulatory variable). Although a variety of effects might conceivably be hypothesized for the grouping variable, one possibility is that costs gravitate toward the average cost level within the category and that the "true" scale effects are distorted

Table 1: The Determinants of Average Operating Costs in New York State, 1975

Independent Variable	*Regression Coefficient*	*Standard Error*	*Beta Coefficient*
	Output Characteristics		
Scale			
Beds1	-0.00183	0.0598	-0.00104
Beds2	-0.0238	0.0269	-0.0465
Beds3	0.00910	0.0101	0.0362
Beds4	0.0146	0.0148	0.0369
Beds5	0.00364	0.00640	0.0164
Patient Age Distribution			
%Under65	-0.0101	0.0807	-0.00641
%Age75to84	0.0114	0.0529	-0.00986
%Over84	-0.0127	0.0448	-0.0164
%Female	0.0290	0.0296	0.0327
Patient Functional Status Variables			
(percentages of patients requiring assistance)			
%Walk	-0.0119	0.0166	-0.0301
%Eat	0.0453***	0.0134	-0.129
%Bathe	0.0319*	0.0193	0.0587
%Breathe	-0.00634	0.00658	-0.0258
%Dress	0.00550	0.0179	-0.0159
%Correspond	0.0229*	0.0127	0.0720
%Medicate	-0.00258	0.0175	-0.00404
%Massage	-0.00200	0.0102	-0.00636
AdmRate	1050***	148	0.332
Service Offerings (0–1)			
Drugs	-0.337	0.599	-0.0156
Phys Therapy	0.211	0.852	0.00705
Occ Therapy	0.770	0.587	0.0364
Speech Therapy	-0.324	0.623	-0.0157
Special Duty	-0.667	1.50	-0.0118
Phys Services	0.935	0.654	0.0458
Other Services	0.520*	0.300	0.0650
Service Intensity			
#Prescriptions	0.00132	0.00284	0.0114
#X-rays	0.341***	0.112	0.0846
Physical Therapy	3.14***	0.976	0.108
Speech Therapy	0.343	6.14	0.00163

Table 1: continued

Independent Variable	Regression Coefficient	Standard Error	Beta Coefficient
Occ Therapy	0.844	1.47	0.0157
Inhal Therapy	14.6**	7.02	0.0535
Radiation Therapy	- 0.933	3.83	-0.000582
Soc Service	1.02**	0.429	0.104
Bedroom Distribution			
%1Bed	0.0316*	0.0158	0.0715
%3Bed	- 0.0165	0.0118	-0.0428
	Input Price Variables		
LPN Wage	0.701***	0.187	0.106
Retail Wage	6.29***	2.31	0.179
PopDen	0.00000865	0.00004	0.0139
	Regulatory-Type Variables		
HRF	- 8.82***	1.47	-0.389
Sponsorship			
Profit	-11.6***	1.44	-5.56
Nonprofit	- 5.64***	1.24	-2.50
Payer Distribution			
%Private	- 0.0364**	0.0174	-0.0733
%Medicare	0.336***	0.0994	0.0940
Survey Certification Ratings			
NI	- 2.40**	1.06	-0.0584
Deferred	0.516	1.09	0.0116
Bed Size Categories			
Size2			
Size3	1.12	0.782	0.0512
Size4			
Size5			
Regions			
Region1	- 1.36	1.27	-0.0389
Region2	- 0.841	1.11	-0.0281
Region3	- 0.696	1.26	-0.0221
Region4	2.88*	1.58	0.0965
Region5	5.07***	1.73	0.151
Region6	4.02**	1.76	0.182

Table 1: continued

Independent Variable	Regression Coefficient	Standard Error	Beta Coefficient
Technology or Production Process Variables			
SNF-HRF	-1.79**	0.706	-0.0745
HRF-SNF	-1.72	1.14	-0.0691
Clinic	-2.79	1.97	-0.0342
%Contract	-0.0651	0.0514	0.0625
Facility Age	-0.0182	0.0130	-0.0626
Indicators of Short-Run Disequilibrium			
Occupancy Segments			
Occ1	0.0442	0.0428	0.0484
Occ2	-0.223	0.179	-0.0745
Occ3	0.267	0.300	0.0507
Occ4	-0.00893	0.102	0.00265
Constant	4.23		
Summary Statistics			
Coefficient of Determination (R^2)	0.768		
Corrected Coefficient of Determination	0.736		
F-Ratio	24.0		

*Significant at 0.10 level
**Significant at 0.05 level
***Significant at 0.01 level

within each group. In any event, it was not clear a priori that scale and regulatory effects could be disentangled. Our effort was exploratory in nature. It quickly became apparent that multicollinearity with respect to nine bed-size variables was a serious problem. We dealt with it by entering all continous (Beds1–Beds5) variables but entering only those categorical (Size2–Size5) variables having a coefficient larger than its own standard error ($t>1$). Examining the results, we find that only one bed-size variable (Size3) enters the relationship. The estimated bed-size relationship is portrayed graphically in Figure 2. We conclude that economies or diseconomies of scale with respect to bed size are minimally important cost factors, although the estimated coefficients indicate very modest economies of scale up to 100 beds and modest diseconomies thereafter.

Occupancy. The results for the occupancy variables (OCC1–OCC4) are more consistent with the usual expectations, although none of the results are significant at the 0.10 level. The model suggests slightly decreasing costs between 80 and 90 percent occupancy, as indicated by a negative coefficient for the OCC2 variable (see Figure 3). The results also indicate increasing costs above 90 percent occupancy; it is not uncommon to find increasing costs as production nears capacity. In any event, the results with respect to occupancy are not dramatic, probably because the capital and other more nearly fixed costs were not included in the AOC dependent variable.

Regulatory-Type Variables. After controlling for differences in patient and service mix, average per diem operating cost is nine dollars less in a Health Related Facility (HRF). For-profit facilities are estimated to cost twelve dollars less than otherwise similar public facilities; and nonprofit facilities are estimated to cost six dollars less than otherwise similar public facilities. Thus, by elimination, nonprofit facilities are estimated to cost six dollars more than otherwise similar for-profit facilities.[16] These results are highly significant.

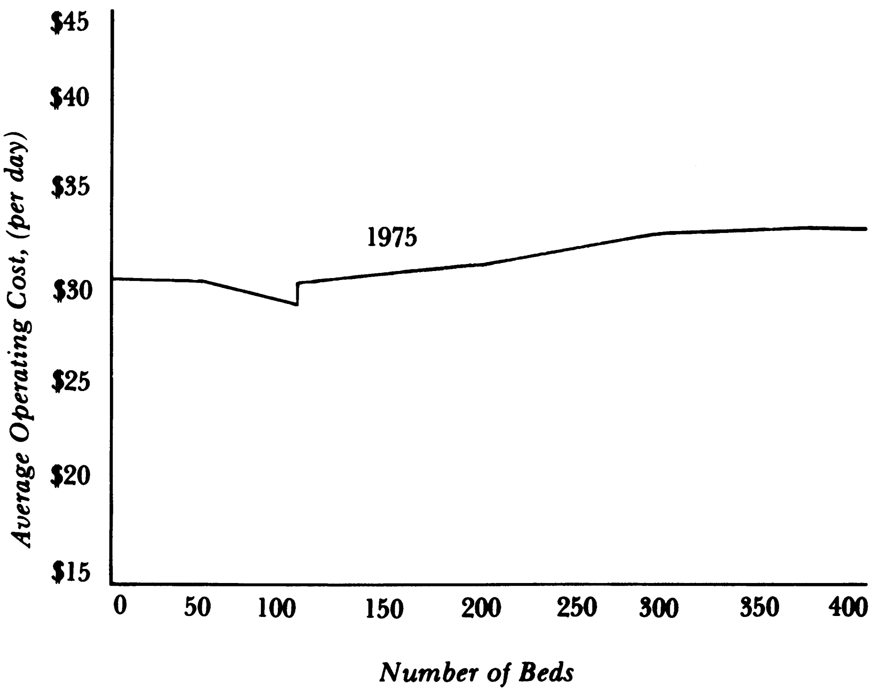
The results for the survey certification variables—NI ("needs improvement") and Deferred—are consistent with expectations. The results indicate the NI facilities cost about $2.40 less per day, probably due to not having incurred the costs of satisfying the certification requirements. The Deferred or rating "deferred" facilities were estimated to cost about $0.50 more.

The model indicates that private patients cost less than otherwise similar Medicaid patients and that Medicare patients cost significantly more than either private or Medicaid patients. The estimates imply that the per diem cost would be $3.64 less for a facility having 100 percent private rather than Medicaid patients, and that the per diem operating cost would be almost $35 higher for a facility with 100 percent Medicare patients. Since we have already controlled for patient characteristics, service mix, and certification difference, it is not clear what differences in care, if any, result from the increased cost of public patients.[17]

Controlling for regional and facility wage levels, there are significant coefficients for several of the regional grouping variables (Region1–Region6). In particular, compared with the West region, the North Metropolitan (Region4), Long Island (Region5), and New York City (Region6) regions cost considerably more—as much as six dollars more per day. Again, it is not clear what differences in care correspond to these cost differences since patient and service mix differences have been controlled for.

Patient Variables. The patient age distribution variables (%Under65, %Age75to84, and %Over84) are insignificant. Functional status and impairment variables apparently reflect relevant age-specific differences.

Figure 2: Bedsize Relationships—1974, 1975 and 1976

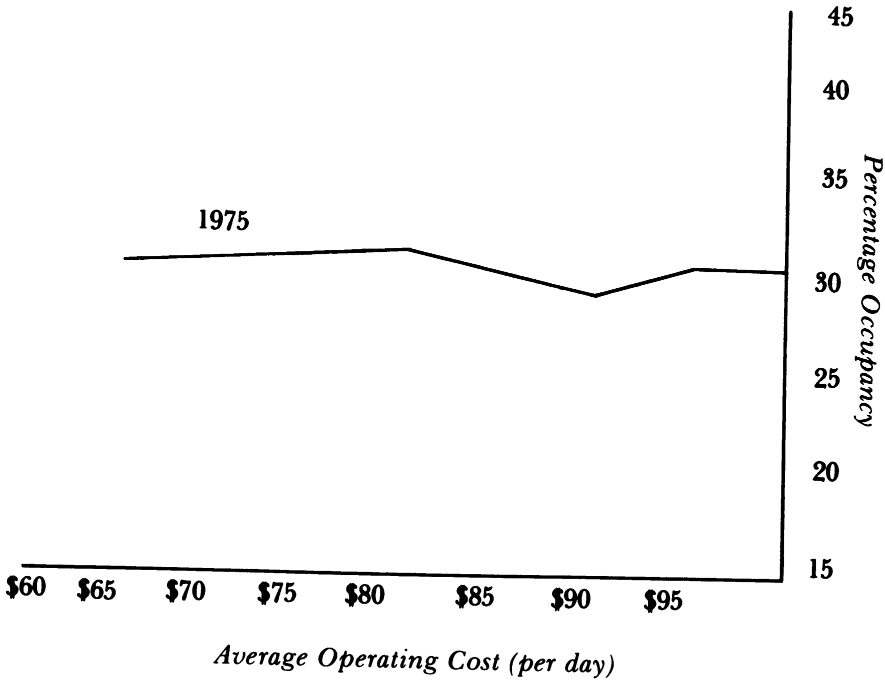


Percent female (%Female) is positively related to cost—as was found by the AMS study—although the coefficient is not significant. The cost per admisison—the coefficient for the ADMRATE variable—is estimated to be $1050.

The model included eight functional status variables (%Walk, %Eat, %Breathe, %Dress, %Correspond, %Medicate, and %Massage). These variables indicate the percentages of patients requiring either partial or full assistance and positive signs were predicted for all of them. Four of the eight variables are positive and all three significant variables (%Eat, %Breathe, and %Correspond) have positive coefficients. Of course, these results may reflect substantial multicollinearity.

Quality Proxies. We had information on seven service intensity variables (#Prescriptions, #X-Rays, Phys Therapy, Speech Therapy, OCC Therapy, Inhal Therapy, Radiation Therapy, and Soc Services) which might be thought to reflect differences in the quality of care. Of course, we anticipate that homes providing greater service intensity will have higher

Figure 3: Occupancy Relationships Estimated from the Complete Model—1974, 1975 and 1976



costs, although multicollinearity may be a problem here as well. Seven of the eight coefficient estimates are positive as predicted. Furthermore, of the four that are significant (#X-Rays, Phys Therapy, Inhal Therapy, and Soc Service), all are positive. We also had additional information on the offering of services and their inclusion in the Medicaid rate (Drugs, Phys Therapy, OCC Therapy, Speech Path, Special Duty, Phys Services, Other Services). Only the coefficient obtained for Other Services was significant; it was positive, as expected. However, the positive coefficients obtained for OCC Therapy and Phys Services are also significant at reasonably high levels, albeit not the usual 0.10 level.

The percentages of patients in rooms having either one, or three or more beds per room—%1Bed and %3Bed—were included as quality measures of a less clinical nature. As expected, one-bed rooms cost more and three-bed rooms cost less, when compared with two-bed rooms.

Price Variables. The average retail wage (Retail Wage) in the county of facility location is used as an index of regional wage levels. The resulting coefficient estimate is positive, as expected, and highly significant. The average LPN wage (LPN Wage) within each facility also gave a significant and positive coefficient, although it is about one-tenth the size of the coefficient for Retail Wage. The county-wide population density (POPDEN), included as a control for systematic price differences in more urban or populated areas, is insignificant.

Technology Variables. Average operating cost is negatively related to facility age (Facility Age). Although the coefficient is insignificant, this result is somewhat unexpected. It had been thought that newer facilities would be more efficient to operate and less costly to maintain. On the other hand, newer facitilies may also support a higher amenity level of care (e.g., air conditioning). The model yields negative but insignificant coefficients with respect to the percentage of services contracted (%Contract) and for facilities maintaining an independent clinic (Clinic). Initial hypotheses were partially supported. Facilities appear to contract for those services that can be obtained more cheaply elsewhere and economies accrue to those facilities having an independent clinic attached. Finally, the results indicate that combined facilities (SNF-HRF and HRF-SNF) cost less, although the result is only significant for an SNF part of a combined facility (i.e., the SNF-HRF variable). Nevertheless, these findings are consistent with the premise that economies accrue to joint SNF and HRF production. If costs were merely shifting between the SNF and HRF parts, opposite signs would have been obtained for the two variables.

Groups of Variables. While the models discussed here give a "good fit," the interpretation of the coefficients is somewhat uncertain in light of the degree of multicollinearity resulting from the large number of independent variables, many of them measuring related factors. We have sought to respond to this problem below by assessing the significance and quantitative importance of entire groups of variables.

Entire groups of variables were tested for statistical significance using a residual F-test. The results are reported in Table 2. The variable groups are significant at the 0.05 level or better if the F-statistic is greater than or equal to the indicated criterion value. The following variable groups are seen to be significant at the 0.05 level: (1) ambulatory status variables, (2) service intensity variables, (3) bedroom distribution variables, (4) price variables, (5) sponsorship variables, (6) certification variables, (7) payor distribution variables, (8) region dummies, and (9) production process variables. The age and sex variables are insignificant. The service offering, bed-size, occupancy, and bed-size and occupancy combined categories are also insignificant. Finally, although the survey certification

Table 2: Residual F-Test Results

Variable Group	F-Statistic	Significance Criterion for .05 Level
Bedsize		
Beds1-Beds5 and Size2-Size5	0.94	1.85
Occupancy—Occ1-Occ4	0.54	2.23
Bedsize and Occupancy	0.74	2.23
Ambulatory Status		
%Walk, %Eat, %Bathe, %Breathe, %Dress, %Correspond, %Medicate, %Massage	4.44	1.96
Age and Sex Distribution		
%Under65, %Age75to84, %Over 84, %Female	0.05	2.30
Service Intensity		
#Prescription, #x-rays, Phys Therapy, Occ Therapy, Inhal Therapy, Rad Therapy, Soc Services	5.39	1.96
Service Offerings		
Drugs, Phys Therapy, Occ Therapy, Speech Therapy, Special Duty, Phys Services, Other Services	1.35	2.03
Bedroom Distribution		
%1Bed and %3Bed	3.98	3.02
Input Prices		
LPN Wage, Retail Wage, Popden	8.54	2.62
Sponsorship		
Profit and Nonprofit	33.65	3.02
Certification—HRF Only	36.00	3.86
Payor Distribution		
%Private, %Medicare	7.55	3.02

Table 2: continued

Variable Group	F-Statistic	Significance Criterion for .05 Level
Survey Certification Rating		
NI and Deferred	2.68	3.02
Technology or Production Process		
Clinic, %Contract, and Facility Age	2.67	2.23
Region—Region2 through Region7	5.74	2.23

variables are not significant at the 0.05 level, they are significant at the less conservative 0.10 level. However, the reader should bear in mind that such results depend substantially upon the specifications of the model within which the variable groups have been tested.[18]

DISCUSSION OF POLICY IMPLICATIONS

This section reviews only those findings that have larger policy significance. First, the analysis indicates that "scale" and occupancy are minimally important to determining operating cost variation. Although increasing returns are indicated between 0 and 50 beds and diminishing returns above 300 beds, these relationships are neither particularly significant nor dramatic. Furthermore, the analysis found no meaningful relationship between cost and occupancy. It appears that regulators have been too much concerned with minimally important scale factors at the expense of other relevant differences that have much more dependable and quantitatively important effects on the cost level.

Second, patient- and service-mix differences do make a difference. The New York State Moreland Commission [6] had concluded from analysis of the substantially limited data available from 1973 cost reports that "patient characteristics and patient needs by and large do not significantly explain variations." Our analysis of the more recent and more exhaustive 1975 cost report data does not support this conclusion. Although age and sex variables are neither statistically significant nor quantitatively important, the patient functional status variables (e.g., percentage of patients requiring assistance with eating) are highly significant as a group and very important in terms of explaining the pattern of nursing home cost differences.[19] Service intensity variables (e.g., number of physical therapy sessions) are similarly significant and important. In addition, several service offering variables were statistically significant.

Third, the sponsorship, certification, and region variables were not only highly significant but also gave comparatively large coefficients. Having estimated one of the most general cost functions to date, it is nevertheless found that nonprofit facilities in New York cost $6.00 more per day than otherwise similar for-profit facilities, that SNFs cost $9.00 more per day than HRFs after controlling for service- and patient-mix differences, and that facilities in the New York City area cost $4.00 to $5.00 more than other facilities in New York State after controlling for regional and even facility-specific wage differentials. The question needs to be asked: What accounts for these differences if not the differences in patient mix, service mix, or input prices measured in this study? Surely, differences of this magnitude cannot continue to be reimbursed without knowing what causes them. If they are merely due to unmeasured differences in the amenity level of care or to managerial inefficiency, then it is not clear that public programs should continue to discriminate among facilities in these categories and pay the differences.

ACKNOWLEDGEMENTS

We thank the Office of Health Systems Management in New York State for making the data for this study available. Mr. Michael Schwartz did most of the programming. We are also indebted to our colleagues on the larger study, Dr. Christine Bishop and Ms. Gail Jensen. Dr. Bishop and Ms. Jensen read and commented on earlier drafts of this paper but, more importantly, helped to shape an overall methodologic approach that was common to all analyses in the larger study. Finally, we wish to thank Dr. Thomas Kickham, the HCFA Project Officer.

NOTES

1. This paper was prepared as part of a much larger study [1] of nursing home costs and the reimbursement alternatives. The larger study also includes econometric cost analyses for the states of Massachusetts and Indiana, as well as single and multiequation cost estimation from the 1973–1974 NCHS national Nursing Home Survey.
2. An SNF is a Skilled Nursing Facility and an HRF is a Health Related Facility, essentially the same as an ICF or Intermediate Care Facility in other states.
3. These variables might once have been used to proxy systematic patient and wage level differences. However, with the genesis of more detailed reporting

requirements—such as those in New York—it becomes unnecessary to rely upon such proxy measures for taking account of relevant differences.

4. Inasmuch as many regulations (e.g., staffing requirements) apply uniformly to all facilities within a state, one does not have any within-state variation, and such regulatory variables must be excluded from a state-specific analysis such as this one. Their effects are reflected in the constant. See [1] for a cross-state analysis that explicitly estimates the consequences of regulatory differences between the states.
5. Methodologically, the accountant's valuation of capital cost is a very imperfect measure of the true economic cost; to a large extent, the reported capital cost more nearly depends upon the age of the facility, the debt/equity structure of financing, the method of depreciation, and the extent of recapitalization due to ownership turnover. The amount of property taxes paid is a similarly unreliable measure and does not meaningfully reflect the differential value of municipal services provided to different facilities. The relative tax burden depends upon the vagaries of property value assessment; and public facilities are tax-exempt. Furthermore, "profit" is simply determined as a residual category and bears a very imperfect relation to actual profitability. Finally, as a practical matter, almost all states reimburse operating and nonoperating expense separately. The issues involved in determining reimbursement for nonoperating expense are altogether different.
6. Bedsize and average daily census correlate so highly ($r^2 = .955$) that a methodological distinction between them is rarely meaningful; regression results are virtually equivalent.
7. Although the final ratings were available we chose not to use them, for both theoretical and empirical reasons. First, the final ratings, as distinguished from the initial ratings, reflect considerable negotiation, including negotiation of improvements or changes to be implemented in the future. Thus, the initial ratings for 1976 should reflect more fairly the situation throughout 1975. This was borne out by preliminary analyses that explored both alternatives. Only 0.8 percent of facilities were designated as needing improvement in the final overall rating. In the initial rating, 7.6 percent of facilities received this designation and 0.2 percent were judged to be altogether unsatisfactory. Thus, we have used the initial ratings in the results reported here.
8. Unfortunately, the data on medication were not usable.
9. Although we might have indicated separate variables for the requirement of partial vs. full assistance, an already serious multicollinearity potential would have been much aggravated. Alternative specifications were pursued in preliminary analysis (e.g., using only the percentage requiring full assistance) and found to be less satisfactory.
10. The average length-of stay was also indicated in preliminary analyses as an alternative measure of patient flow. However, the admission rate variable (AdmRate) gave vastly superior statistical performance.
11. A further service intensity variable, indicating the proportion of patients receiving activity therapy sessions per patient, was omitted because of its collinearity ($r^2 = 0.88$) with Soc Service. The annual number of tests per patient was also omitted for analogous reasons.
12. Admittedly, this is not a very satisfactory proxy for the price of capital. In retrospect, the median value of owner-occupied housing or median residential rent—available by county from the Census of Housing—might have been a

better choice. Alternatively, inasmuch as capital-related expense was excluded from the dependent variable, we could also have used a measure of the capital stock itself such as assets per bed. (Of course, the accounting valuation of capital is a highly imperfect measure of actual worth.) In the absence of a more satisfactory proxy for either capital or the price of capital, we must assume that capital and labor are not substitutable in producing nursing home care. Where others also make this assumption [3,p.50], the proposition has never been tested. To the extent that capital effects are vintage, i.e., related to differences in nursing home design and construction that vary over time, they are proxied by the facility age variable.

13. To take an example, if actual occupancy were 93 percent, then we would have that OCC1 equals 80, OCC2 equals 10, OCC3 equals 3, and OCC4 equals 0.
14. Assuming that the incidence of nonreporting is random, this procedure does not bias the coefficient estimates. However, if all right-hand-side variables were estimated, there would be no gain in efficiency. See [5], pp. 336–344.
15. Similar relationships, not reported here, were also estimated using 1974 and 1976 data. However, the information available was somewhat different in those years. For example, only very limited data were available from the 1974 reports, and extensive patient impairment data were also available for 1976. Because of such problems we did not pursue pooled cross-section/time-series analysis; the common variable list was too small. On balance, the 1975 data were more complete. Many of the coefficient estimates from analysis of the 1974 and 1976 data were surprisingly similar across years despite differences in model specification.
16. The 1976 model (not reported here) found that the nonprofit/for-profit differential dropped abruptly to four dollars in that year when sponsorship was eliminated as a grouping category for reimbursement purposes.
17. In the 1976 model, which is not reported here, we also included variables measuring patient impairment along each of the following dimensions: (1) alertness, (2) sight, (3) hearing, (4) speech, and (5) communications. The results imply that, ceteris paribus, the per diem cost would have been $6.18 less in a facility having all private patients and about $28 more in one having all Medicare patients. Even so, one may argue that we have not controlled adequately for patient condition. In particular, data were not available on mentation and continence.
18. A residual F-test was also conducted on the patient impairment variables in the 1976 model. These variables were not found to be significant ($F = 1.04$, $C_{0.06} = 2.23$).
19. Several patient impairment variables were also significant in the 1976 analysis which is not reported here.

REFERENCES

1. Birnbaum H et al. *Public Pricing of Nursing Care.* 1982. Cambridge MA: Abt Books.
2. Applied Management Sciences, Inc. Analysis of incentive reimbursement systems for health care and LTC services provided to the elderly and long-term

disabled. August 1976. Final report to the Office of the Assistant Secretary for Planning and Evaluation, DHEW (Contract No. HEW 100-76-0029), Rockville MD.

3. Bishop CE. Nursing home cost studies and reimbursement issues. *Health Care Financing Review* (Spring 1980) 1:4.
4. Fuller, WA. Grafted polynomials as approximating functions. *Australian Journal of Agricultural Economics* (June 1969).
5. Kmenta J. *Elements of Econometrics*. 1971. New York: MacMillan.
6. Moreland Commission. Reimbursing Operating Costs: Dollars Without Sense. 1976. Report of the New York State Moreland Commission on Nursing Homes and Residential Facilities. Albany NY. p. 54.