YOU GET WHAT YOU PAY FOR:
Using Input Prices To Validate Quality Of Care Measures

Steven D. Pizer, Ph.D.
Man Wang, M.S.
Catherine Comstock, M.P.H.*

November 26, 2004

ABSTRACT

Economic theory predicts that improvement in quality of care requires commitment of resources in a production function. We use this prediction to validate seven quantitative measures of nursing home quality. Measures based on data external to the nursing home were positively and significantly associated with input prices (nurses' wages) in regression models controlling for output volume and case mix, and measures based on facility-reported data were not. This implies that quality measures based on externally reported data might be less subject to measurement error than those relying on facility-reported resident assessments (like those chosen by the VA and by Medicare).

Dr. Pizer is Health Economist, Department of Veterans Affairs and Assistant Professor, Boston University School of Public Health. Ms. Wang and Ms. Comstock are with Boston University and the Department of Veterans Affairs, respectively.

A previous version of this paper was presented to the Allied Social Science Associations Annual Convention, January 5, 2004, in San Diego, California.

JEL Code: I18

Key Words: Cost and Quality, Nursing Homes, Quality Measurement

Correspondence to:
Steven D. Pizer, Health Care Financing and Economics, Boston VA Healthcare System
Mail Stop 152H, 150 South Huntington Ave., Boston, MA 02130
Phone: (617) 232-9500 x6061; Fax: (617) 278-4511; E-mail: pizer@bu.edu

YOU GET WHAT YOU PAY FOR:
Using Input Prices To Validate Quality Of Care Measures

November 26, 2004

ABSTRACT

Economic theory predicts that improvement in quality of care requires commitment of resources in a production function. We use this prediction to validate seven quantitative measures of nursing home quality. Measures based on data external to the nursing home were positively and significantly associated with input prices (nurses' wages) in regression models controlling for output volume and case mix, and measures based on facility-reported data were not. This implies that quality measures based on externally reported data might be less subject to measurement error than those relying on facility-reported resident assessments (like those chosen by the VA and by Medicare).

INTRODUCTION

Health care payors, providers, and consumers have long sought broadly accepted quantitative measures of the quality of health care services as a means of improving management and purchasing. Health policy makers have responded by supporting research that identifies new potential measures and refines existing ones. This has been particularly true with respect to nursing homes because nursing home residents are less able than others to advocate on their own behalf, and because budget pressures in recent years have threatened to force compromises in quality (IOM, 1986; GAO, 2002a; Angelelli et al., 2002).

In 2002, hoping to improve quality through increased competition and public scrutiny, and following the example set by several state Medicaid programs, the Centers for Medicare and Medicaid Services (CMS) announced plans to make nursing home quality measures available to the public over the Internet (Harris and Clauser, 2002). The announcement of this nursing home quality initiative certainly accelerated research efforts, but may have rushed the decision to rely on some measures rather than others (GAO, 2002b; Manard, 2002). As was the case with state efforts, the federal initiative sifted through a variety of potential measures, eventually settling on a subset to be publicized (Berg et al., 2002). However, the question remains: Were the best quality measures selected?

This paper appeals to economic theory and available data to begin to answer this question. Theory generally characterizes quality as an output, and as such, production of quality requires inputs (Cohen and Spector, 1996; Gertler and Waldman, 1992; Nyman, 1985). Increases in input prices, on average, ought to force reductions in quality, and decreases in input prices ought to have the opposite effect. Perhaps surprisingly, the assorted quality measures currently in use in nursing homes have never been evaluated to determine which ones meet this basic expectation. Those that prove to be related to input prices could be prioritized to reflect the relative strength of this relationship, guiding oversight efforts toward areas most likely to suffer when budget pressures mount. Those that prove to be unrelated to input prices may help to identify data quality problems requiring attention.

We use Veterans Health Administration (VA) data to examine seven quality measures developed for use in both VA and community nursing homes. Two measures rely primarily on data collected externally to the nursing home—either when VA patients receive hospital care or when they die—and the other five measures rely on patient assessment data, collected by VA nursing home staff on a semi-annual schedule. We find that the externally reported measures were significantly associated with input prices (nurses' wages) in regression models controlling for output volume and case mix and the measures based on facility-reported data were not. This implies that facility-reported quality measures (like those chosen by the VA and by Medicare) may suffer from more measurement error than measures that rely on externally reported data.

The remainder of this paper is organized as follows. In the next section we briefly review the economic theory of production with endogenous quality as applied in the literature to nursing homes, and then adapt it to the VA setting. Next we discuss the data files we assembled, our sample, risk-adjustment approach, and statistical specification. Finally, we present results, and provide some discussion of the implications of our work.

THEORETICAL BACKGROUND

The economic theory of production has been applied to the nursing home industry to study the effects of changes in payment policy on quality of care. The shift from cost reimbursement to prospective payment, holding average reimbursements constant, has been shown by Cohen and Spector (1996) to lead a profit-maximizing nursing facility to reduce quality. Furthermore, if prospective payment were imposed uniformly by all payors and the average level of payment were reduced while holding other variables constant, quality of care would decline even more. This is not, however, how prospective payment is implemented. As pointed out by Nyman (1985, 1988a, 1988b) and Cohen and Spector (1996), if capacity is constrained and payments are reduced by public payors (Medicare or Medicaid) only, then quality can improve as providers attempt to increase average payments by attracting greater numbers of private-pay residents.

Even if capacity is not constrained, substantial reductions in payment may or may not produce important declines in quality, depending on returns to scale in quality (see Gertler and Waldman, 1992). This is defined as the percentage change in cost brought about by one percentage point change in quality. Increasing returns to scale in quality (cost changes less than quality in percentage terms) suggests that large changes in quality might be expected in response to changes in payment levels. By contrast, decreasing returns to scale in quality suggests the opposite.

A Simple Model

The following model, loosely based on Gertler and Waldman (1992), helps to establish the empirical challenges we face when modeling private sector nursing homes.

Assume that nursing homes attempt to maximize profits as characterized by Equation (1),

$$\Pi = PY^p + R(Y - Y^p) - C \qquad (1)$$

where Y and Y^p denote total resident-days and private-pay resident-days, respectively; P and R represent the price charged to private-pay and public-pay residents, respectively; and C is the cost of providing care. The number of resident-days accounted for by private-pay residents is a function of the price, P, and the quality of care, Q, as well as some other factors that affect demand for private-pay care (like local per capita income and the availability of home care), denoted by Z. This relationship is summarized in Equation (2).

$$Y^p = Y^p(P, Q, Z) \qquad (2)$$

Furthermore, the cost of care is a function of the total number of resident-days, Y, the quality of care, Q, and the cost of inputs (primarily wages), W. Equation (3) reflects this relationship.

$$C = C(Y, Q, W) \qquad (3)$$

Assuming that facilities are at or very near capacity, they maximize profits by choosing P and Q, generating the first order conditions given by Equations (4a) and (4b),

$$Y_P + PY^p_P - RY^p_P = 0 \qquad (4a)$$

$$PY^p_Q - RY^p_Q - C_Q = 0 \qquad (4b)$$

where subscripts denote partial derivatives with respect to P and Q. By the implicit function theorem, Equations (4a) and (4b) could be solved to give the equilibrium values of P and Q as functions of the exogenous variables (Y, W, R, and Z), allowing Y^p to be calculated afterwards.

This model permits changes to R to be evaluated in terms of their effects on P, Q, C, and Y^p. Unfortunately, P is unmeasured in most nursing home data and any results would depend critically on the functional forms chosen to characterize $Y^p(.)$ and C(.). $Y^p(.)$ is particularly complex because it involves the interplay between facility behavior and private demand for nursing home care, a relationship affected by numerous unobservable factors. Nevertheless, returns to scale in quality can be measured without having to estimate the entire system. To do this we focus attention on the last term in Equation (4b), C_Q. This is the partial derivative of cost with respect to quality and when multiplied by the ratio of quality to cost (Q/C) gives returns to scale in quality. The other two terms taken together constitute the total effect of a quality change on facility revenue, so Equation (4b) indicates that quality will be set such that the marginal increase (or loss) in revenue is equal to the marginal increase (or saving) in cost.

Although the quality effects on revenue are complex and difficult to measure, we can estimate the quality effect on cost using Equation (3) without measuring P or imposing a functional form on $Y^p(.)$. All we need is a functional form for C(.) and a means of measuring C and Q.

Unfortunately, Q is an endogenous variable in Equation (3). To see this, consider an unexpected increase in the exogenous variable W. Higher W directly causes an increase in C, changing the value of C_Q in Equation (4) and inducing new equilibrium values of P and Q. Thus, causation runs from W to C to Q. By contrast, consider a shift in the exogenous variable Z (e.g., supply of alternatives). Demand from private pay patients changes, affecting P and Q and resulting in a new value of C. This time causation runs from Z to Q to C. Because of this endogeneity, instruments for Q would have to be found to estimate the returns to scale in quality from Equation (3). Gertler and Waldman (1994) suggested using elements of Z as instruments (including per capita income and the proportion of the local population over age 65), but these have proven to be weak in practice (Mukamel and Spector, 2000).

An additional problem with this approach is that in recent years private sector nursing homes have no longer been capacity constrained (Bishop, 1999). Since Y is no longer fixed, another endogenous variable is introduced into the model, making estimation even more difficult.

A Model for VA Facilities

Some of the difficulties with the private nursing home model are ameliorated in a study of public facilities. In particular, there is only one source of revenue, instead of a mix of public and private sources, and public facilities have been consistently capacity constrained in recent years.

Each year Congress appropriates a budget for VA patient care, and funds are distributed among 21 regional networks according to a formula based on the numbers of patients treated in each network in the previous year. Network directors then distribute funds among medical centers and center directors distribute funds among hospital units and outpatient clinics. In principle, any funds not spent at the end of the year revert back to the Treasury.

Loosely following Newhouse (1970) and Sloan (2000), assume that nursing home managers maximize a utility function given by Equation (5), where the notation is consistent with the private sector model above.

$$U = U(Y, Q, \Pi) \tag{5}$$

Because any profits would have to be returned to the Treasury, this utility function is effectively maximized subject to the zero-profits constraint given by Equation (6).

$$\Pi = RY - C = 0 \tag{6}$$

Cost is described by the same function as in the private sector, given by Equation (3). Two more constraints apply to the problem. First, capacity cannot be expanded easily because capital spending must be centrally authorized and funded. As a result, the number of patient days (Y) cannot exceed a capacity limit. Second, because VA patients could always choose private sector facilities instead, VA nursing homes must offer quality of care above some minimum or they would lose their patients to private alternatives. These constraints are given by Equations (7) and (8), respectively.

$$Y \leq \overline{Y} \tag{7}$$

$$Q \geq Q_{min} \tag{8}$$

Assuming that $U_Y > U_Q$ in the relevant range, two solutions are possible. If R is greater than some R*, then the capacity constraint in Equation (7) will be binding and the quality constraint in Equation (8) will not. If R is less than R*, then the quality constraint will be binding and the capacity constraint will not. These equilibria are described by Equations (9a) and (9b).

$$C(\overline{Y}, Q, W) = R\overline{Y} \qquad \text{if } R>R^* \tag{9a}$$

$$C(Y, Q_{min}, W) = RY \quad \text{if } R<R^* \tag{9b}$$

The VA pays for nursing home care in three settings: VA nursing homes, state veterans' homes, and community nursing homes. Veterans typically prefer VA nursing homes because of lower out-of-pocket costs and more comprehensive care, turning to community nursing homes when VA beds are not available (GAO, 1996; GAO, 2001). Veterans' preferences imply two things for our model. First, VA nursing homes are nearly always at capacity, and second, fluctuations in budget authority may be principally absorbed by adjustments in VA payments to state veterans' homes and community nursing homes, an issue we will return to at the end of this section.

Given that VA nursing homes are capacity constrained, Equation (9a) can be solved for the equilibrium value of Q as a function of exogenous variables $\overline{Y}$, W, and R, as shown in Equation (10).

$$Q = Q(\overline{Y}, W, R) \tag{10}$$

Equation (10) cannot be estimated directly because, at the nursing home level, R could be endogenous. Exogenous decisions by Congress and the distribution formula determine budgets at the network level, but network and medical center directors might respond to quality problems at a particular nursing home by shifting resources from other network facilities, thereby creating the appearance of a negative relationship between quality and funding. To correct this problem, an exogenous instrument for R is needed. Fortunately, a suitable instrument can be found in the preliminary network-level distribution of the annual appropriation. This distribution is a function of the congressional appropriation and the distribution formula, neither of which are sensitive to quality variations at individual nursing homes.[1] We produce a nursing home-level instrument by scaling the preliminary network-level allocation by the fraction historically received by each facility. This is a powerful method in part because VA nursing home budget allocations are very stable from year to year.

The very stability that produces a powerful instrument, however, is evocative of another problem alluded to earlier. If hospital directors absorb fluctuations in budget authority mainly by adjusting VA payments to state veterans' homes and community nursing homes, then we should not expect to observe a substantial relationship between R and Q in Equation (10), regardless of the quality of our instrument. A recent VA report (HCFE, 2003) indicates that, not only do hospital directors insulate their nursing homes from budget fluctuations, but they are required to do so by law. Consequently, our primary empirical focus is on the relationship between W and Q in Equation (10), controlling for any variations in R that may occur. Our principal hypothesis is that increasing wages for nursing staff in the local labor market in conjunction with an inflexible budget allocation will force local managers to substitute away from labor or hire lower quality workers, resulting in lower quality of care. Likewise, declining local wages would have the opposite effect.

DATA AND SAMPLE

The Medicare Nursing Home Quality Initiative selected five quality measures for long term residents: % with loss in ADL functioning, % with pressure ulcers, % with pain, % in physical restraints, and % with infections. All of these measures are constructed from facility-reported resident assessment data. VA data collected during our study period did not support these measures, so we selected three that were similar and four that represented alternatives, for a total of seven: change in ADL score, incidence of pressure ulcers, prevalence of urinary tract infections, prevalence of dehydration, change in behavior score, rate of potentially preventable hospitalization, and mortality rate. The first five relied primarily on facility-reported resident assessments and the last two relied primarily on data collected from sources external to the nursing home.

To construct these outcome measures, resident-level data were extracted and merged from several VA administrative data sets. In general, health and functional status information from a baseline point in time were used to risk-adjust an outcome measured up to six months later (the follow-up point). Because all residents in VA nursing homes were assessed on a semi-annual schedule, once in April and again in October (as well as upon admission or transfer), data from these scheduled assessment points can be arrayed in a time series and treated as a sequence of representative samples which can be analyzed for evidence of improvement or deterioration in quality of care over time. Although it would no doubt be better from a monitoring point of view to have more frequent assessment points, the fact that all nursing home residents were assessed at regular intervals is enough to provide a reasonable assurance of representativeness. Because our study period was three years long (October 1997-October 2000), we had six sub-periods, each anchored by a semi-annual assessment date.

Resident assessment data for this analysis were drawn from the VA Patient Assessment File (PAF), which contains patient assessments collected in April and October of each year.[2] All VA nursing home residents in those months were assessed using a standard form (the Patient Assessment Instrument). Complete data are readily available from April 1986 to October 2000, after which a transition began to a new form, the Minimum Data Set or MDS. In addition to the PAF, the Patient Admission and Transfer (PAT) file contains patient assessments in the same format, generated by patients who were assessed either at admission or transfer (from acute care into the nursing home) sometime between PAF assessment dates.

For risk adjustment purposes, we supplemented assessment data from the PAF and PAT with diagnosis codes extracted from Bed Section records contained in the Patient Treatment File (PTF) and Extended Care File (ECF). All diagnosis codes were selected from records dating between 365 days prior to 14 days after a baseline assessment date. Thus all diagnoses recorded during inpatient encounters up to one year prior to assessment were linked to the baseline assessment record. These diagnosis codes were grouped into broad categories and specific disease variables following Rosen et al. (2000).

For the five outcomes that relied on facility-reported assessment data, each outcome measure was constructed by relating risk adjustment information from one patient assessment (the baseline assessment) to outcome information from a subsequent patient assessment (the follow-up assessment). Because representativeness is important, the universe of PAF assessments was selected to define the sample. Following Rosen et al. (2000), each PAF assessment (April or October) was considered to be a follow-up and was paired with the immediately preceding assessment for that patient, whether from the PAF or the PAT. Thus, the longest interval between baseline and follow-up was about 7 months (e.g., if the baseline were done early in April and the follow-up were done late in the following October). The shortest interval permitted was 14 days and pairs of assessments with shorter inter-assessment intervals were excluded from the analysis.

Because the goal of this analysis was to evaluate quality of care in nursing homes, it was critical that patients whose outcomes were considered were actually in the nursing home during the interval between assessments. For this reason, follow-up assessments were paired with the immediately preceding assessment even though this meant that the inter-assessment interval was not standardized (e.g., to six months).[3] Using a standard inter-assessment interval would have resulted in some patients having one or more assessments between baseline and follow-up, implying that some fraction of the interval must have been spent outside the nursing home (e.g., in an acute care ward).

Our measure of cost was derived from the VA's Cost Distribution Report (CDR).[4] We extracted costs associated with staff assigned through the nursing service to the nursing home care unit. These personnel included registered nurses, licensed practical nurses, nurses' aides, and a few administrative support staff. To instrument for cost we obtained the annual network-level budget allocations produced by the Veterans Equitable Resource Allocation (VERA) System. These allocations were downloaded from the VA's Allocation Resource Center (ARC), accessible through the VA intranet.

Our measures of wages were obtained from the Outgoing Rotation Groups of the Current Population Survey (CPS-ORG).[5] We tabulated mean weekly wages by geographic area for registered nurses (RN's), licensed practical nurses (LPN's), and certified nurses' aides (CNA's), for 1998, 1999, and 2000. Geographic areas were defined to be metropolitan statistical areas (MSA's) corresponding to VA medical centers, provided sample sizes were large enough. In some cases we combined MSA's or used state boundaries instead to increase sample size.

Sample

Approximately 7% of the assessments in the PAF file could not be matched to any preceding assessment, indicating that these patients were admitted on a date close enough to the semi-annual assessment date that a separate admission assessment was not done for them. Of the remaining assessments, approximately 10% were excluded because the inter-assessment interval was less than 14 days. The remaining sample of 57,858 follow-up assessments was distributed across the six time periods as shown in Table 1.

The preventable hospitalization and mortality samples differed because they did not rely on paired assessments. Instead, the universe of PAF assessments (April and October) was used for baseline information, and survival or preventable hospitalization status over the following six months was determined from records generated outside the nursing home. Consequently, it was not necessary to exclude those with inter-assessment intervals of less than 14 days, resulting in a larger initial sample of 64,529.

CONSTRUCTION OF QUALITY MEASURES

We selected seven measures of quality, choosing some that have been developed specifically for VA nursing homes and some that we adapted from their application to community nursing homes or other uses. We begin by discussing those that rely primarily on data generated by patient interactions outside the nursing home (i.e., inpatient admissions and dates of death), and then move on to those that rely primarily on facility-reported data.

Analysis of risk-adjusted mortality requires as a first step the accurate determination of when and if each patient died during the study period. For patients who died in the hospital, dates of death were extracted from VA inpatient and extended care records. Additional dates of death are recorded in the BIRLS file for VA enrollees and in Medicare enrollment records for Medicare beneficiaries. The dates of death from these two sources occasionally conflict, so we have prioritized the Medicare dates in prior work (CHQOER, 2003) due to the fact that these dates corresponded more consistently with inpatient records. Rather than rely exclusively on date of death records, it is advisable to confirm wherever possible that those not listed as deceased were still alive (Berlowitz et al., 1997, CHQOER, 2003). We did this by locating the most recent record of VA utilization and inferring that the patient was alive at least until that date. Data sets checked for such evidence included PTF, ECF, outpatient clinic, pharmacy, and fee basis files. Using these methods, we were able to confirm survival status for 98.5% of the cases included in the study.

The next outcome of interest is the rate of potentially preventable hospitalization. Potentially preventable hospitalizations were defined according to the AHRQ definitions (AHRQ, 2001) for 13 distinct types of hospitalizations (AHRQ also defined 3 types that reflect pediatric diagnoses, but these were not appropriate for our study population). These types are identified by ICD-9-CM principal diagnosis codes and by procedure codes. In the case of Medicare-financed hospitalizations, each admission was evaluated to determine whether it matched criteria characterizing preventable hospitalizations. Each admission thus identified was flagged to reflect which type it was (e.g., asthma, diabetes, COPD) and the date of admission was recorded. Because VA nursing home residents can move from the nursing home care unit to an acute care unit and back without recording a new admission, we used a sub-admission unit called a bedsection stay as the equivalent of an admission.[6] Thus, bedsection stays that meet the preventable hospitalization criteria, based on diagnosis and procedure codes, were flagged and the corresponding dates were recorded.[7]

Following Berlowitz et al. (1997, 1996), pressure ulcer incidence (development of new pressure ulcers) was defined to be no ulcer or a stage 1 ulcer at baseline and stage 2 or worse at follow-up. Dehydration and urinary tract infection prevalence were defined as presence of the condition at follow-up, regardless of baseline dehydration or UTI status, respectively. This approach is consistent with the "sentinel event" interpretation[8] suggested by Zimmerman et al. (1995), with the modification that we risk-adjust these measures as suggested by more recent literature (e.g. Mukamel, 1997).

Functional status and behavior problems were both measured on scales ranging from 3 to 15. The functional status scale was constructed, following Rosen et al. (2000), as the modified sum of activity of daily living (ADL) ratings for eating, transferring, and toileting.[9] The behavior problem scale was constructed similarly as the sum of ratings for physically aggressive, disruptive, and verbally disruptive behavior (Porell et al., 1998). Functional status and behavior changes were both calculated by subtracting baseline scores from follow-up scores.

For the sake of consistency and ease of exposition, we constructed each outcome variable so that larger positive numbers signify less favorable outcomes. Thus, for example, a change in a patient's functional status score of +3 indicates more dependence in activities of daily living from baseline to follow-up. Similarly, dichotomous variables were coded with zero representing the absence of the condition and one representing its presence (e.g., dehydrated, new pressure ulcers, urinary tract infection, died within six months). Means and standard deviations of the unadjusted outcome variables are presented in Table 2.

Resident-level Risk Adjustment

To be consistent with the bulk of the risk adjustment literature, we applied resident-level regression techniques to adjust each outcome measure for differences in resident risk at baseline. This type of risk adjustment is critical when attempting to compare quality between facilities that may have very different resident populations. A facility that specializes in the care of residents suffering from Alzheimers' disease, for example, would be expected to have higher rates of behavior problems than a facility that specializes in post-acute care. If the two facilities were compared without accounting for differences in their baseline risk, the possibly erroneous conclusion would be drawn that the Alzheimers' facility provides lower quality care.

Because the analysis does not attempt to compare facilities, the importance of risk adjustment ought to be reduced. However, if the average case mix of individual VA nursing homes changed in a way that was correlated with costs or input prices, then risk adjustment would make a noticeable difference and, nevertheless, is a reasonable precaution.

Risk Adjustment Variables

Where possible, this study applies risk adjustment models that were developed for use with PAF data, including the functional status change model developed by Rosen et al. (2000) and the pressure ulcer and mortality models developed by Berlowitz et al. (1996, 1997). Because PAF-based models were not available in the literature for dehydration, urinary tract infection, or behavior problems, we combined elements from the other models into a hybrid model, which we estimated separately for each of these three outcomes.[10] In selecting variables for the hybrid model, we chose to err on the side of inclusiveness, incorporating variables that were shown in the literature to be useful for risk adjusting pressure ulcers, mortality, or functional status change unless they were clearly redundant (see Appendix A for a complete list).

Depending on which outcome was to be risk adjusted, data from the same semi-annual (PAF) assessment was used either as baseline information (externally reported outcomes) or as follow-up information (facility-reported outcomes). For the five facility-reported outcomes, the dependent variables were outcomes measured from PAF assessments (treated as follow-ups) and the independent variables were drawn from the immediately preceding assessment (PAF or PAT treated as baseline). These included baseline values of the outcome as well as potentially associated health and functional status variables supplemented by diagnoses from inpatient and outpatient records. In the mortality models, the dependent variable was six-month mortality drawn from several sources external to the nursing home and the independent variables were baseline health and functional status from PAF assessments and diagnoses, both individually and combined into the Charlson index—a measure of co-morbidity (Charlson et al., 1987; Deyo, Cherkin, and Ciol, 1992). In the preventable hospitalization models, the dependent variable was whether the resident had one or more preventable hospitalizations over the six months following baseline (PAF) assessment and the independent variables were the same as in the hybrid models, with the addition of lagged preventable hospitalization.

Thus, starting from the same semiannual (PAF) assessment, the facility-reported outcomes were constructed by looking backwards for baseline information, and the externally reported outcomes were constructed by looking forwards for six months for deaths or hospitalizations. Appendix A provides definitions for all the independent variables and Appendix B reports coefficients used for each model. Table 2 reports risk-adjusted means and standard deviations for each outcome.

STATISTICAL MODEL

Equation (10) related quality of care to output level, input prices (primarily wages), and resources expended. This relationship is expressed in Equation (11), imposing a linear functional form and substituting the variables in our data for their theoretical counterparts.

$$Q_{it} = \beta_0 + \beta_1 Cost_{it} + \beta_2 Volume_{it} + \beta_3 CaseMix_{it} + \beta_4' Wages_{it} + \beta_5' Year_t + \varepsilon_{it} \qquad (11)$$

The variable Q denotes each of the seven risk-adjusted quality measures; Cost represents total nursing cost; Volume reflects the number of residents in each facility; CaseMix is a variable constructed to reflect the average degree of medical complexity of each facility's residents (predicted mortality in this case); Wages is a vector of wage rates for registered nurses (RN's); licensed practical nurses (LPN's), and certified nurses' aides (CNA's); and Year is a vector of year effects. The subscripts i and t index facilities and years, respectively.

There were two statistical challenges involved in estimating this relationship: 1) the quality measures (Q) were heteroskedastic, and 2) the cost variable (Cost) was endogenous, so instruments were needed.

Heteroskedasticity arises because the quality measures that constitute the dependent variables in Equation (11) were all constructed as facility-level mean residuals from individual-level regression models. Because different facilities had different numbers of residents, these means were calculated from samples of different sizes, leading to heteroskedasticity. To correct this problem, we clustered on facility when calculating standard errors for each model.

Cost was endogenous in Equation (11) for the same reasons that the corresponding variable (R) was endogenous in Equation (10). If a particular facility developed quality problems during the year, hospital or regional managers might have responded by redirecting resources from other programs to the troubled facility. Thus, poor quality could cause high cost. To account for this we employed the instrumental variables (IV) technique, selecting as instruments the preliminary network-level budget authorizations, scaled by the ratio of facility expenditure to network-level budgets from 1997 (the year prior to the study period). Although the IV estimation was performed as one step, we also estimated an illustrative model that demonstrates the power of these instruments (Table 3). The model as a whole achieved an R-square of 0.57, and the scaled budget authorizations were highly significant, producing a t-statistic of 28.

RESULTS

As a first step, we estimated Equation (11) by ordinary least squares (OLS), intentionally neglecting the endogeneity of cost. Results indicate that these models had low predictive power, with R-squared statistics ranging from 0.02 to 0.07 (Table 4). Few of the independent variables had statistically significant effects, but these were generally in the direction predicted by theory. Higher RN wages were positively associated with preventable hospitalizations and CNA wages were positively associated with mortality, although LPN wages were negatively associated with ADL Change. Higher values of the case mix variable were also positively associated, as predicted, with three types of quality problems: dehydration, ADL change, and behavior change. Perhaps surprisingly, the cost and volume variables had small and insignificant effects on all quality measures.

When we performed the estimation by IV, overall predictive power declined (as is typical with IV), but the statistically significant results hardly changed. Of the effects discussed above, only the case mix effect on dehydration dropped below the level of significance (Table 5). Preventable hospitalizations were significantly associated with RN wages (coefficient = 0.004), mortality was significantly associated with CNA wages (0.02), and ADL change was negatively associated with LPN wages (-0.07). Thus, if the typical facility faced a 10% increase in weekly RN wages (from $626 to $689), preventable hospitalizations could be expected to increase by 0.25 percentage points (0.004*63), 3% of the unadjusted rate. Similarly, a 10% increase in CNA wages (from $302 to $332) would be associated with an increase in mortality of 0.6 percentage points, or 4% of the unadjusted rate. ADL change and behavior change were both positively associated with case mix (coefficients = 274 and 110, respectively), implying that a 10% increase in predicted mortality (from 0.174 to 0.191) would be associated with a 0.5 point increase in ADL score (274*0.017/100)[11] and a 0.02 point increase in behavior score.

The fact that our results changed so little with the IV approach is consistent with the lack of influence of the cost variable in the model. This is consistent with the idea that VA hospital directors absorb fluctuations in budget authority by adjusting payments to state veterans' homes and community nursing homes before reducing budget authority for the VA nursing home care unit.

DISCUSSION

In this article we apply the economic theory of production to empirically validate quality of care measures in VA nursing homes. Some might object to this premise, arguing that quality improvements do not always require expenditure of additional resources, particularly in a public sector organization like VA. We offer three points in response to such concerns. First, we do not deny that some quality improvements can be had for free, but we firmly believe that systematic quality improvement across the VA would require additional resources. This belief equates to an assumption that production is occurring at or near the efficiency frontier, an assumption that may meet with objections when applied to the public sector. This brings us to our second point, which is that VA does not operate in the same organizational environment as private sector institutions. Our approach does not require that VA facilities are as productive as their private sector counterparts (although they may be, see Nugent et al., 2004), only that VA managers are doing about as well as they can given their budgets and the regulatory regime under which they operate. Finally, we wish to note that although we believe resources are required for systematic quality improvement, this assumption is not necessary for our approach to be valuable. If some dimensions of quality can be improved without the application of resources, our approach will not identify them, focusing instead on those dimensions that are vulnerable to compromise when budget pressures mount. Because the effect of budget pressure on quality is a constant concern for quality regulators, we believe this focus is appropriate.

We find significant positive relationships between input prices (nurses' wages) and quality as measured by risk-adjusted preventable hospitalization and mortality rates.

These effects are consistent with theory predicting that higher market wages for nurses will lead facility administrators to reduce quality as long as the level of quality is higher than the minimum quality constraint implied by non-VA competition. These results provide some basis for preferring preventable hospitalization and mortality rates to the other measures of nursing home quality, a preference reinforced by the fact that these two quality measures were the only ones relying on externally reported outcomes. The relative strength of the externally reported outcomes is consistent with the hypothesis that facility-reported outcomes can be expected to exhibit greater measurement error because poor quality facilities may not be as vigilant about data quality as their high-quality counterparts.

Militating against this ranking to some degree were our results on the effects of case mix, which was positively associated with ADL change and behavior change, as predicted by theory. Theory indicates that more medically complex resident populations will exhibit higher rates of unfavorable outcomes, even after adjusting for risk, because our risk adjustment models are imperfect. Also consistent with imperfect risk adjustment, we find that the ADL change variable was negatively associated with LPN wages, a result that could arise from the fact that facilities specializing in post-acute rehabilitation tend to be located in cities with relatively high wages. These facilities typically exhibit improvement in ADL status among their patients, hence the observed association between high wages and ADL improvement. These results are harder to interpret than the wage effects because the relationships among our outcomes, our risk adjustment models, and our case mix variable are more complex and potentially ambiguous than those between input prices and quality.

In summary, we found input price effects that support a preference for the externally reported quality measures, preventable hospitalization and mortality. Among the facility-reported quality measures, the case mix effects supported some preference for behavior change and ADL change, although these effects could have been due to weak risk adjustment. Dehydration, pressure ulcer incidence, and UTI prevalence all failed to exhibit in the IV models any relationships predicted by theory (although dehydration was related to case mix in the OLS estimates).

At a minimum, further research is needed to determine whether this pattern will be found with newer resident assessment instruments such as the Minimum Data Set Version 2.0 (MDS), currently in use in Medicare- and Medicaid-certified facilities and recently adopted by the VA. If the relative strength of externally based measures is borne out with MDS data, then the current exclusive reliance on facility-reported quality indicators in the federal Nursing Home Quality Initiative (Harris and Clauser, 2002) should be reconsidered. Furthermore, this research supports attempts to track and improve the accuracy of resident assessments. The Centers for Medicare and Medicaid Services (CMS) has adopted software standards to improve accuracy and has undertaken research projects intended to identify elements of the assessment that are particularly vulnerable to measurement error (Manard, 2002).

Several limitations should be kept in mind when considering the results of this study. First, the sources of both cost data (the CDR) and patient assessment data (the PAF and PAT) are now obsolete. The CDR has been superseded by a new accounting system and the MDS 2.0 instrument has replaced the PAF, starting in FY2001. It is possible, therefore, that a future analysis, exploiting the superior qualities of these data systems, might find different results. Second, we did not attempt to develop the best possible risk adjustment model for each outcome, so it is possible that better risk adjustment could lead to better performance for some measures. Finally, this research was conducted on a sample of VA nursing home residents, a population known to differ from residents of community nursing homes in some important ways (e.g., most VA residents are men and most community nursing home residents are women). Consequently, similar analysis should be performed on a sample of community nursing home residents before any conclusions are applied to non-VA nursing home policy.

NOTES

[*] The authors wish to thank Ann Hendricks, Dan Berlowitz, Amy Rosen, and Chapin White as well as seminar participants at Boston University and session attendees at the ASSA Annual Convention for helpful comments. This research was supported by the Health Services Research and Development Service of the Department of Veterans Affairs, Management Consultation Grant MRR 00-018. All opinions expressed are those of the authors and do not necessarily reflect the position of the Health Services Research and Development Service or the Department of Veterans Affairs.

[1] For a technical description of the VHA distribution formula, see the ARC Technical Manual – VERA 2002, available from the VA Allocation Resource Center.

[2] Summary information about the collection, use, and maintenance of PAF data is recorded in the VHA Corporate Databases Monograph, available from the Veterans Affairs Information Resource Center (VIReC): http://www.virec.research.med.va.gov/library.

[3] We controlled for differences in inter-assessment intervals by including this interval in our risk adjustment models (see Appendix A).

[4] The quality of CDR data is discussed in Swindle et al. (1996), and Nugent et al. (2003) suggest adjustments that researchers should make to improve comparability with private sector payment categories.

[5] We chose to use market wages rather than VA wages because we were concerned that the federal wage scale might not accurately reflect local labor market conditions, despite recent reforms in this direction. If local wages increase and federal wages do not adjust, then federal managers will have to either "game" the wage scale or make do with lower quality workers. In either case, local market wages would better reflect the situation than federal wages would.

[6] A new bedsection stay record is created whenever a patient is "admitted" to a VA acute care unit (either from the community, from another institution, or from some other unit of the VA medical center (like the nursing home). Because this method of record keeping differs from what is done in non-VA hospitals, preventable hospitalization rates might not be directly comparable between VA and non-VA patients. Because our sample consists exclusively of VA patients, this does not present a problem for our analysis.

[7] Because our study population consisted of VA nursing home residents, most acute care hospitalizations occurred in VA hospitals (94%). We tested whether the inclusion of Medicare-financed hospitalizations made a substantial difference for this population and found that it did not (the overall average rate increased from 9.1% to 9.7%). Nevertheless, to be comprehensive, we included Medicare-financed hospitalizations in our rates of preventable hospitalization.

[8] The term "sentinel event" refers to a rare and troubling situation that arguably should never occur. Therefore, its occurrence signals potentially serious quality problems, even without risk adjustment.

[9] Note that this definition (Rosen's) differs from some others in that only three ADLs are used. We followed this standard because it was developed specifically for VA data.

[10] Note that we did not attempt to construct the best possible risk adjustment model for each outcome. Our more limited goal was to apply a model that would eliminate well-known sources of case mix variation and facilitate comparison among quality measures.

[11] We scaled up the quality measures by 100 to improve the stability of regression calculations. This produces results in percentage points instead of probabilities for most measures, but results for ADL change and behavior change should be scaled back to the original metric.