Certificate-of-Need Regulation and Entry: Evidence from the Dialysis Industry*

JON M. FORD
DAVID L. KASERMAN

Auburn University
Auburn, Alabama

I. Introduction

Certificate-of-need (CON) regulation is widely used in the health care industry.¹ The primary alleged purpose of this regulatory tool is to reduce industry costs by preventing "unnecessary duplication of facilities."² While specific CON regulations vary from state to state, virtually all require new firms planning to enter the industry and incumbent firms planning an expansion of productive capacity to submit an application in which the applicant must demonstrate: (1) a market demand (or "need") for the incremental output, investment, or new service being proposed, and (2) the inability or unwillingness of existing firms to meet that demand with facilities already in place. Moreover, incumbent firms are offered the opportunity to formally intervene during the CON review process to express their opposition to the proposed entry or expansion plans.

Economists have long been skeptical of this form of regulation. At least three fundamental reasons underlie this skepticism. First, private investors are likely to have vastly superior information to that held by regulators on the need for new capacity. These investors are much more familiar with industry conditions than regulators, and they are placing their own money at risk by entering and/or expanding. Second, given the obvious incentive of existing firms to oppose virtually any entry, expansion of capacity, or introduction of new services by competitors and the fact that this policy provides an open forum for such opposition, the likelihood that CON regulation actually serves the interests of consumers by fostering lower industry costs is remote. And third, to the extent that CON regulation is effective in reducing net investment in the industry, the economic effect is to shift the supply curve of the affected service back to the left. Since most medical services are thought to exhibit inelastic demand (due to the general unavailability

*The authors thank Bob Ekelund, John Mayo, John Jackson, Steve Caudill, and an anonymous referee for helpful comments on a previous draft. The usual caveat applies.

1. See Fine and Super [6], Graham [8], and Coyte [4]. Lanning, Morrisey, and Ohsfeldt [13, 151] provide a brief synopsis of the history of CON regulation of hospitals:

Although some states implemented CON programs in the 1960s, most states implemented CON programs in the mid 1970s, due at least in part to the federal Health Planning and Development Act of 1974. By 1980, all states except Louisiana had enacted a CON review program. However, federal funding for CON review programs was substantially cut in 1981–82, and the Health Planning Act was repealed outright in 1986. By 1987, 13 states had ended CON review for hospitals.

2. Other alleged purposes are identified in Sloan [18].
of substitutes and the high frequency of third-party payments), the effect of such supply shifts is to raise both equilibrium price and total expenditures on the affected service, which is precisely opposite of the stated objective. Despite these criticisms, however, CON regulation remains a pervasive force in the health care industry.

The economic criticisms outlined above assume that CON regulation represents a binding constraint on capacity expansion decisions. There has been some recent debate in the literature, however, about whether CON regulation is, in fact, effective in reducing net new investment in the industry. Some authors have argued that the CON review process does not prevent new firms from entering or existing firms from expanding but merely requires these firms to justify their capacity expansion plans to regulators. Any investments warranted by market conditions, they argue, are generally approved. Thus, there is some doubt as to whether CON regulation represents a binding constraint on investment in the affected industry.

In this paper, we investigate this issue by examining the impact of CON regulation on entry into the dialysis industry over the decade of the 1980s. This industry grew substantially during this period of time. On December 31, 1980, there were 1,041 dialysis clinics with 12,329 stations in the U.S. By December 31, 1989, there were 1,830 clinics with 23,654 stations [20]. In part, this growth is attributable to a 1972 amendment to the Social Security Act which authorizes the federal government to pay 80 percent of the cost of treatment (by either dialysis or kidney transplantation) of all citizens suffering from renal failure. The End Stage Renal Disease (ESRD) program, which is operated under Medicare, has grown from $229 million in its initial year (serving 11,000 patients) [7] to $3.7 billion in 1988 (serving 110,000 patients) [23]. Such increases in funding have provided strong incentives for entry into this industry. And the presence of these incentives, in conjunction with changes in CON regulation, provides an ideal experimental situation in which to measure the impact of such regulation on observed entry.

An attractive feature of our study is that we are able to utilize two alternative measures of entry. Consequently, our results are important not only for policy decisions regarding CON regulation in this and other health-related industries but also for evaluating the empirical performance of the different entry measures used. Given the recent re-emergence of the perceived role of entry (or potential competition) as an important disciplining force affecting market behavior, our findings should be of widespread interest.

The paper is organized as follows. First, we describe two alternative measures of entry used in this study. These measures are made possible by the very detailed accounting of both firms and capacity reported in the dialysis industry. Second, we specify a simple empirical model of the

3. The idea that holding down investment in an industry can reduce costs apparently originates in some confusion about fixed costs versus total costs and long-run costs versus short-run costs. Advocates of CON regulation correctly argue that restricting entry and capacity expansion will lead to increased utilization of existing facilities (which will cause movement down the short-run average fixed cost curve). There is no guarantee or even likelihood, however, that increased utilization will lead to lower average total costs. Moreover, there is even less likelihood that restrictions on capacity expansion will cause long-run average (or total) costs to be minimized at any given level of output.

4. See Sloan and Steinwald [19], Sloan [17], and Mayo and McFarland [14]. Lanning, Morrisey, and Ohsfeld [13, 144] state that: “The general consensus in the literature is that CON review has had little or no effect on hospital costs or expenditures . . . .”

5. In this study, we focus on independent for-profit clinics performing hemodialysis. Hemodialysis is, by far, the predominant form of dialysis, accounting for 98.9 percent of the total number of patients. See U.S. Department of Health and Human Services [20]. Among the independent (i.e., non-hospital based) dialysis clinics, for-profit facilities account for approximately 83 percent of the total number of firms. See U.S. Department of Health and Human Services [21].

determinants of observed entry into this industry. Next, we describe our data and present our empirical results. These results indicate that CON regulation has significantly retarded new firm entry and total capacity expansion in this industry, thereby restricting supply and fostering increased levels of industry concentration. Finally, we summarize our findings and conclude the paper.

II. Measures of Entry

Most prior empirical studies of the entry process have measured entry by the net change in the number of firms in the industry over some specified period of time [3; 5; 15]. This measure, however, is widely recognized as being deficient in two important respects. First, it does not account for the size of the new firms that have entered the industry nor the size or any incumbent firms that have exited. And second, a simple count of the net change in the number of firms also fails to reflect expansion or contraction of capacity by incumbent firms already in the market. For both reasons, this traditional entry measure falls short of the general concept of entry as an overall expansion of productive capacity in an industry.

Here, we utilize two different measures of entry into the dialysis industry. First, we employ the traditional measure—the annual net change in the number of dialysis clinics in each state in each year in our sample (1982 through 1989). This measure (E1it) is included in order to compare the results obtained with our alternative entry measure.

Second, we employ the annual net change in the total number of dialysis stations (i.e., machines) in each state in each year. This measure (E2it) reflects both new firm entry (including size of firms and number of firms) and incumbent firm capacity expansion. Thus, it represents a broader measure of entry than the traditional one in that it reflects capacity additions by existing firms. At the same time, it is a more refined measure, because it accounts for the size of both the new firms coming into the market and existing firms leaving.

III. Model Specification

In this section, we explain the theoretical justification for including the various exogenous variables incorporated in our empirical model of entry into the dialysis industry. The variable that is of primary interest for this study is our measure of the presence of CON regulation in each state. Accordingly, CONit is a binary variable equal to one if CON regulation of dialysis clinics was in place in state i in year t and is zero otherwise.4

If CON regulation has been effective in curtailing entry and expansion in this industry, the

7. Prior to the widespread adoption of CON programs, states had been encouraged to implement planning programs to prevent "unnecessary capital expenditures" under Section 1122 of P.L. 92-603, a 1972 amendment to the Social Security Act. Later, in 1974, the National Health Planning and Resource Development Act (P.L. 93-641) specifically required states to develop CON review programs or lose their Public Health Service funds [22]. By 1980, these CON programs had largely replaced the 1122 planning programs; and, as Joskow [11] argues, the remaining 1122 programs are redundant in states with a CON program. Consequently, a separate variable for 1122 programs is not included.

8. Information regarding the status of CON regulation in each state over our sample period was obtained through a survey of state health planning agencies administered by the authors. In this survey, we asked the following question: "Did your state have certificate-of-need regulation applying to dialysis facilities for the following years: 1981, 1982, . . . 1989." Thus, we are able to determine which states had CON programs that specifically applied to dialysis clinics for each year in our sample.
coefficient of this variable should be negative and significant in our entry equation.\(^9\) If, on the other hand, CON regulation represents a non-binding constraint on new investment in the dialysis industry, the coefficient of this variable should be insignificant in this equation. Thus, our empirical results should help resolve the issue of whether this regulatory tool is effective in reducing entry and/or expansion in the dialysis industry.

In those states where CON regulation is applied to dialysis industry investments (i.e., where \(CON_{it} = 1\)), the stringency of the regulatory constraint may vary depending upon the threshold investment levels required to trigger a CON review. These thresholds vary from a low of zero in several states to a high of $1,000,000 in Alaska and North Carolina.\(^{10}\) To account for these threshold investment levels, we define a binary variable, \(T_i\), that is equal to one for those states specifying a zero threshold and is zero otherwise. This variable is then interacted with \(CON_{it}\) to provide a binary variable that is equal to one only when (a) the state has a CON review program applicable to dialysis clinics, and (b) that program applies to all investments, regardless of their magnitude. This interaction variable is incorporated along with \(CON_{it}\) to measure the additional impact on entry that CON regulation is likely to exert where thresholds are zero. Together, \(CON_{it}\) and \(CON_{it} \cdot T_i\) will reflect the overall effect of this type of regulation on entry.

Because the size of the investment required to open a new dialysis clinic is much greater than that required to expand an existing clinic, \(CON_{it} \cdot T_i\) is more likely to have a significant effect where entry is measured as a net change in the number of dialysis machines in place (\(E2_{it}\)). New dialysis machines can generally be purchased for less than $15,000. Consequently, in those states where CON regulation is subject to a positive threshold (i.e., where \(T_i = 0\)), capacity expansion in existing clinics can often escape the review process. New clinics, however, generally require an investment that exceeds most, if not all, threshold levels. Thus, the size of the threshold level is likely to be irrelevant to the entry of new firms, and the interaction variable is unlikely to be significant where entry is measured as a net change in the number of firms (\(E1_{it}\)). Consequently, \(CON_{it} \cdot T_i\) is included in the \(E2_{it}\) (number of machines) equation but not in \(E1_{it}\) (number of firms).

Additional variables are included in the model to control for other important influences on entry into this industry. Six such variables are incorporated.\(^{11}\) First, we include registered nurses’ wages (\(WAGE_{it}\)) in our model.\(^{12}\) The dialysis business is relatively labor intensive—labor costs may account for as much as 70–75 percent of the total costs of operating a clinic. Generally, one registered nurse, licensed practitioner nurse, or technician is required for every two or three

---

9. Mayo and McFarland [14] measure the stringency of state CON programs by the ratio of denials to total applications. Such a measure may or may not be superior to the one used here, because it fails to reflect the "discouraged applicant" effect—that is, projects that never get proposed due to an expectation that they will not be approved. We are unable to employ such a measure here due to data limitations.

10. See American Health Planning Association [1]. These data are available only for a single year. Assuming that the individual states with CON review programs have not varied their threshold levels over time, however, this limitation should not affect our results.

11. We are unable to incorporate a direct measure of industry profitability or price in our model. Data on profitability are simply unavailable. Price (or reimbursement) data do exist, but the Health Care Financing Administration sets the price in this industry. Moreover, these reimbursement rates do not vary substantially from state to state, and what little variation does exist is intended to reflect interregional differences in nurses' wages, which is included as a variable in our model. The correlation coefficient between price and nurses' wages is 0.68. In addition, these reimbursement rates have been changed only twice over the decade of the 1980s. Thus, the available price data exhibit insufficient variation in the sample and are correlated with another variable included in the model.

12. Data for this variable and two others introduced below are available only for one year. For clarity, we shall drop the time subscript for these three variables.
patients being dialyzed. Patients must be monitored fairly closely during their treatments to control the amount of fluid being removed and to respond to various problems that commonly arise during treatments (e.g., cramps, nausea, and hypotension). Thus, \( WAGE_i \) reflects the observed variation in one of the principal determinants of costs in this industry. Consequently, it should reflect the variation in profitability of the firms in this industry. And, because \( WAGE_i \) is positively related to costs, it is negatively related to profitability. Thus, we expect a negative sign on the coefficient of this variable in our entry equation.

Second, we include the percent of the state’s population that is black \( (PB_i) \). The incidence of renal failure (and, therefore, dialysis) is relatively high among the black population. Moreover, due to improved funding, information, transportation facilities, and the increasing availability of dialysis clinics in rural areas, many more blacks began receiving dialysis treatments during the decade of the 1980s. Consequently, as a determinant of demand, the percent of the state’s population that is black should have a positive effect on entry into the dialysis industry over this period of time. Thus, we anticipate a positive sign on the coefficient of \( PB_i \).

Third, renal failure is more prevalent among older citizens. Specifically, the incidence of kidney disease is significantly higher among those over forty-five years of age [21]. Consequently, we include the percent of each state’s population that is forty-five years old or older, \( PA_i \), as an additional determinant of the demand for dialysis services. We expect the coefficient of \( PA_i \) to obtain a positive sign in our entry equation.

Fourth, we include average per capita income in state \( i \) in year \( t \), \( PCI_{it} \). Theoretically, we should expect a positive coefficient for this variable as higher incomes may increase demand. Although dialysis treatment is covered under the End Stage Renal Disease program of the Health Care Financing Administration (HCFA), patients’ incomes may still exert a positive impact on demand for two reasons. First, HCFA does not provide full coverage of the costs of dialysis. Rather, they reimburse clinics 80 percent of the estimated costs of dialyzing patients. The patient’s ability to pick up the remaining 20 percent depends upon their income, either through direct payment or through insurance coverage. Thus, clinic profitability is likely to be positively influenced by patient’s incomes. Second, higher income individuals are relatively more inclined to seek out whatever medical care is needed. Thus, for a given population and a given incidence of renal failure, higher incomes are likely to generate a higher demand for dialysis services. Therefore, both profitability and the level of demand are likely to be positively associated with \( PCI_{it} \).

Fifth, the overall demand for dialysis service is likely to be positively affected by the population of the state, given some average probability of renal failure. Thus, we include the population in state \( i \) in year \( t \), \( POP_{it} \), in our entry equation, and we anticipate a positive sign for the coefficient of this variable.

And sixth, as a result of interstate migration and divergent rates of population growth, states have experienced different changes in population over the decade of the 1980s. Consequently, the growth of demand for dialysis service is likely to have differed from state to state. To account for the influence of this factor, we incorporate the annual change in each state’s population in each year over the sample period, \( CPOP_{it} \), in our entry equation. Since population growth should also increase the demand for dialysis services, ceteris paribus, we hypothesize a positive sign for the coefficient of this variable.

Our empirical specification of the two entry equations, then, is given by

\[
E1 = \beta_0 + \beta_1 CON + \beta_2 WAGE + \beta_3 PB + \beta_4 PA + \beta_5 PCI + \beta_6 POP + \beta_7 CPOP + \epsilon_1
\]  

(1)
Table 1. Variable Names, Definitions, and Data Sources

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_{1it}$</td>
<td>Net change in the number of dialysis clinics in state $i$ in year $t$.</td>
<td>[1]</td>
</tr>
<tr>
<td>$E_{2it}$</td>
<td>Net change in the number of dialysis stations (i.e., machines) for state $i$ in year $t$.</td>
<td>[1]</td>
</tr>
<tr>
<td>$CON_{it}$</td>
<td>Binary variable equal to 1 if CON regulation of dialysis clinics was in place in state $i$ in year $t$.</td>
<td>[2]</td>
</tr>
<tr>
<td>$T_i$</td>
<td>Binary variable equal to 1 for those states specifying a zero CON threshold in 1990.</td>
<td>[5]</td>
</tr>
<tr>
<td>$WAGE_i$</td>
<td>Annual salary of newly licensed RNs in state $i$ in 1990.</td>
<td>[3]</td>
</tr>
<tr>
<td>$PB_i$</td>
<td>Percent of population that is black in state $i$ in 1980.</td>
<td>[4]</td>
</tr>
<tr>
<td>$PA_i$</td>
<td>Percent of population over age 45 in state $i$ in 1980.</td>
<td>[4]</td>
</tr>
<tr>
<td>$PCI_{it}$</td>
<td>Per capita income in state $i$ in year $t$.</td>
<td>[4]</td>
</tr>
<tr>
<td>$POP_{it}$</td>
<td>Population in state $i$ in year $t$.</td>
<td>[4]</td>
</tr>
<tr>
<td>$CPOP_{it}$</td>
<td>Change in population in state $i$ in year $t$.</td>
<td>[4]</td>
</tr>
</tbody>
</table>

[2] Survey of fifty states by the authors.

and

$$E2 = \delta_0 + \delta_1 CON + \delta_2 CON \cdot T + \delta_3 WAGE + \delta_4 PB + \delta_5 PA$$

$$+ \delta_6 PCI + \delta_7 POP + \delta_8 CPOP + \epsilon_2,$$  \hspace{1cm} (2)

where we have dropped the subscripts denoting states and time, and $\epsilon_1$ and $\epsilon_2$ are random disturbance terms. Variable names, definitions, and data sources are provided in Table 1.

IV. Data and Empirical Results

Because our sample contains annual observations on individual states over the 1982–1989 time period, we have panel data. Entry is measured as annual changes in either the number of firms ($E1$) or the number of dialysis machines ($E2$). Consequently, the data for these two variables actually begin in 1981. All fifty states are included in the sample, and we have eight annual observations. Thus, our sample contains 400 observations. Data for four of our exogenous variables ($T$, $WAGE$, $PB$, and $PA$) are available only for a single year. Thus, we are assuming that CON thresholds, nurses’ wages, percent black, and percent over forty-five years of age do not vary over our sample period. This assumption seems reasonable in light of the fact that these factors tend to vary considerably more across states than across time.

Because these data contain both time-series and cross-sectional observations, efficient estimation requires use of a panel estimator. Here, we utilize the Parks [16] estimation technique, which assumes heteroskedasticity, a first-order autoregressive error structure, and contemporaneous correlation between cross sections [12, 512–514]. Given these assumptions, the covariance matrix for the vector of random errors can be estimated by a two-stage procedure, and the model’s
<table>
<thead>
<tr>
<th>Variable</th>
<th>E1</th>
<th>E2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON</td>
<td>-1.490*</td>
<td>-3.369***</td>
</tr>
<tr>
<td></td>
<td>(-3.28)</td>
<td>(-1.88)</td>
</tr>
<tr>
<td>CON-T</td>
<td>—</td>
<td>-29.254*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-2.55)</td>
</tr>
<tr>
<td>WAGE</td>
<td>-0.196*</td>
<td>0.370</td>
</tr>
<tr>
<td></td>
<td>(-4.00)</td>
<td>(0.24)</td>
</tr>
<tr>
<td>PB</td>
<td>0.075*</td>
<td>0.964*</td>
</tr>
<tr>
<td></td>
<td>(3.05)</td>
<td>(3.28)</td>
</tr>
<tr>
<td>PA</td>
<td>0.036***</td>
<td>2.252*</td>
</tr>
<tr>
<td></td>
<td>(1.62)</td>
<td>(3.41)</td>
</tr>
<tr>
<td>PCI</td>
<td>0.175*</td>
<td>1.496**</td>
</tr>
<tr>
<td></td>
<td>(2.27)</td>
<td>(2.28)</td>
</tr>
<tr>
<td>POP</td>
<td>0.115</td>
<td>1.816**</td>
</tr>
<tr>
<td></td>
<td>(1.44)</td>
<td>(2.17)</td>
</tr>
<tr>
<td>CPOP</td>
<td>0.012*</td>
<td>0.105*</td>
</tr>
<tr>
<td></td>
<td>(7.95)</td>
<td>(5.49)</td>
</tr>
<tr>
<td>Intercept</td>
<td>3.520*</td>
<td>-74.228***</td>
</tr>
<tr>
<td></td>
<td>(2.84)</td>
<td>(-1.76)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.42</td>
<td>0.36</td>
</tr>
</tbody>
</table>

a. $t$-statistics are in parentheses under each coefficient estimate.

* = significant at the .01 level.

** = significant at the .05 level.

*** = significant at the .10 level.

parameters can then be estimated with generalized least squares. These estimates are unbiased, consistent, and asymptotically efficient. The results of this estimation are reported in Table II.

Overall, these results are quite encouraging. Virtually all of the coefficients attain the hypothesized signs, and all but one are statistically significant in each equation. Moreover, the $R^2$s indicate a reasonably high degree of explanatory power for pooled time-series, cross-sectional data.

Turning to the individual coefficient estimates, we find that the presence of CON regulation has had a significant negative impact on both the entry of new firms and the expansion of capacity in this industry over our sample period. The CON variable exhibits a negative and significant coefficient in both of our entry equations. Moreover, when entry is measured as a change in the total number of machines in the industry (E2), we find that having a zero threshold for CON review reduces entry even further.13 Thus, both the presence and the stringency of the CON program reduces entry into the dialysis industry.

Turning to the remaining coefficient estimates, we find mixed results for nurses' wages, WAGE. The coefficient of this variable is negative and significant in the firm entry equation (E1)

13. The program providing the Parks estimates would not run when CON*T was included in the E1 entry equation due to collinearity problems. As noted above, however, there are theoretical reasons to believe that CON review thresholds should have no impact on new firm entry because of the investment required for such entry. Moreover, when the E1 equation is estimated with OLS with the interaction variable included, the coefficient of this variable is insignificant, as hypothesized. The results of that estimation are available from the authors upon request.
but is positive and insignificant in the capacity equation \((E2)\). This result could be due to substitutability between nurses and dialysis machines in the production of dialysis services. One can easily substitute nurses for machines simply by keeping the clinic open longer hours. Given such substitutability in the production process, an increase in nurses' wages will increase the demand for machines at a given output.

The coefficient associated with the percent of a state's population that is black, \(PB\), is positive and significant in both entry equations. Thus, increases in the demand for dialysis services caused by variations across states in racial composition have led to new entry and expansion in this industry. Similarly, the percent of the population that is forty-five years old or older, \(PA\), is also found to exert a positive and significant effect on entry whether \(E1\) or \(E2\) is employed as the entry measure.

The coefficient associated with per capita income, \(PCI\), is also positive and significant in both equations. Thus, despite the funding provided by the End Stage Renal Disease Program, higher income areas have attracted greater entry by dialysis clinics. The greater profits available where more patients are either covered by insurance or are financially able to pay the additional twenty percent not covered by the federal program have attracted more clinics and greater capacity in this industry.

Finally, both population and the change in population over the sample period (both demand-side variables) have also exerted a positive and generally significant effect on entry (although \(POP\) is not significant in the \(E1\) equation). Therefore, while CON regulation appears to have constrained new investment in the dialysis industry below what it would have been in the absence of such regulation, it has not completely curtailed the ability of the industry to respond to demand growth.

V. Conclusion

The evidence presented here demonstrates that CON regulation has provided an effective constraint on entry and expansion in the dialysis industry over the decade of the 1980s. It has retarded the growth of new capacity by both new and incumbent firms, as well as growth in the number of firms, thereby contributing to reduced capacity and increased levels of concentration in this industry.

Prior research [9] has shown that heightened levels of industry concentration in dialysis markets leads to an overall deterioration in the quality of care provided as firms with market power attempt to increase profits by lowering costs in the face of fixed (regulated) prices. Additional evidence also suggests that declining quality has contributed to increased patient mortality in this industry [10].

The results presented here suggest that CON regulation has contributed to this increasingly serious quality problem. By maintaining unnecessarily high levels of industry concentration and by restricting supply, CON regulation of the dialysis industry has sustained the monopoly power of incumbent clinics and, thereby, provided the wherewithal to increase profits by reducing service quality. Thus, CON regulation has promoted the interests of incumbent suppliers to the detriment of consumers (patients).
References


