Revisiting Rogowski and Newhouse on the indirect costs of teaching: a note on functional form and retransformation in Medicare’s payment formulas

Kathleen Dalton a,*, Edward C. Norton b

a Cecil G. Sheps Center for Health Services Research, CB# 7590 725 Airport Road, Chapel Hill, NC 27599-7590, USA
b Department of Health Policy and Administration School of Public Health University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Received 1 November 1999; received in revised form 1 May 2000; accepted 26 May 2000

Abstract

In 1992 Rogowski and Newhouse identified errors in functional form and retransformation in the econometric model that underlies Medicare’s payments to teaching hospitals. We re-estimate their model and expand on their work, with data from the following decade. We find: (1) the functional form imposed by Health Care Financing Administration’s original specification of the teaching variable is supported by the data; (2) there is no evidence of a threshold effect when the teaching intensity variable is appropriately specified; (3) there is no longer evidence of heteroscedasticity across teaching hospital types, consequently there is no need to incorporate re-transformation factors into the payment formula. We attribute the differences in our findings to secular changes in the hospital industry and improvements in variable measurement. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Hospital costs; Teaching hospitals; Prospective payment system (PPS); Retransformation

* Corresponding author. Tel.: +1-919-966-7957; fax: +1-919-966-5764.
E-mail address: kathleen_dalton@unc.edu (K. Dalton).
1. Introduction

In 1992 Rogowski and Newhouse published a comprehensive assessment of the regression model developed by the Health Care Financing Administrations (HCFA), that serves as the basis for Medicare’s current system of payment adjustments to hospitals participating in graduate medical education (Rogowski and Newhouse, 1992). The purpose of this note is to determine if their findings still apply to cost data collected over 10 years later, after a period of substantial structural change in the hospital industry. This is an appropriate time to review earlier work, because the Medicare Payment Advisory Commission is currently reviewing the empirical evidence and policy implications behind Medicare’s medical education adjustments in order to recommend changes to Congress (MedPAC, 1999). The indirect medical education (IME) formulas governing these payments affect the distribution of nearly US$5 billion of US$78 billion now paid to short-stay hospitals, and they are sensitive for political as well as budgetary reasons. Over time, IME payments have acquired a policy justification in addition to their empirical basis, in that the adjustment has intentionally been allowed to exceed estimated cost differentials in order to fulfill a policy objective of directing more Medicare funds to teaching hospitals. Resolution of the specification issues in the empirical model remains important, however, in order to identify how much of the payment adjustment reflects true product differentiation and how much should be viewed as an educational subsidy.

1.1. Background

HCFA’s original hospital cost model as described by Pettengill and Vertrees (1982) and Lave (1985), was a log-linear specification where the logarithm of Medicare operating cost per case was estimated as a function of teaching intensity, with controls for case mix, hospital size, regional wage variation and metropolitan area size. The analysis was conducted on hospital-level arithmetic means, using accounting data from federal fiscal year 1981. The ordinary least squares estimation was run with no second-order terms, imposing an assumption of a diminishing marginal effect of changes in teaching intensity on unit cost, with a single teaching coefficient computed across the full sample of hospitals.

Although the specific formula has changed over time, coefficients from this regression have been used to adjust the prices paid by Medicare to teaching hospitals since the start of the Prospective Payment System (PPS). In the enabling legislation (PL 97-248, in 1982) the coefficient on the teaching variable was incorporated into the PPS price structure as a linear adjustment based on the ratio of interns and residents to staffed beds. The formula was revised after 2 years to reflect the diminishing marginal effect implied by the model’s logarithmic form. A later version of the model, constructed by the Congressional Budget Office, also
restricted the control variables and their parameters to conform to the PPS payment formula. The teaching coefficient from this restricted version was adopted as part of PL 99-272 in 1985, and has remained in effect ever since. For reasons related to policy concerns rather than empirical evidence, the teaching adjustment has also been deliberately increased to levels in excess of the measured cost differential, by applying a multiplier to the adjustment that has ranged from 2.0 (in the 1982 legislation) to 1.35 (mandated for 2002, by the Balanced Budget Refinement Act of 1999).

After its initial introduction into legislation, HCFA’s model was the subject of a moderate amount of attention in the health economics literature (Anderson and Lave 1986; Welch 1987; Thorpe 1988; Sheingold 1990). The econometric analysis was criticized primarily for bias attributable to omitted location variables and other operating characteristics, for the theoretical inconsistency of including hospital size in the estimation while failing to include it in the payment formulas, and for failure to recognize a difference in the strength of the teaching effect between large and small teaching programs. Thorpe also speculated that HCFA’s specification of the teaching variable (as the log of one plus the ratio of residents to staffed beds) introduced additional bias into the estimation insofar as the true relationship is assumed to be between cost per case and the ratio itself, without the constant.

Rogowski and Newhouse (1992) systematically assessed these and additional problems in HCFA’s original model, using data from discharges in federal fiscal year 1984. They conducted regression diagnostics to assess heteroscedasticity and omitted variable bias, specified an alternative teaching variable (computed as the natural log of the sum of the resident-to-bed ratio plus 0.0001) and generalized the log-linear assumption by constructing spline functions on their revised teaching variable to allow for multiple slope coefficients. They addressed additional technical shortcomings of the model by adding weights to account for estimation on group means and incorporating retransformation corrections in the error term. They examined differences in the expected teaching effects that were derived from equations with and without selected covariates that had been suggested by other investigators, including the variable for number of beds, Medicaid utilization, and additional indicators for regional location and ownership. They confirmed earlier findings of a threshold level, below which the teaching coefficient was significantly smaller and/or non-significant. Above that threshold they found that their revised formulation of the teaching variable did not produce a substantially different estimate of teaching hospital cost differentials than those produced by HCFA’s specification, but under their specification the teaching coefficient was less sensitive to the inclusion of other covariates. This led Rogowski and Newhouse to conclude that some of what had been identified by other investigators as omitted variable bias was, instead, the result of incorrect functional form. They also found that predicted teaching effects were significantly altered when smearing estimates were included in the computations of expected cost differentials across hospital groups.
Rogowski and Newhouse closed, however, with a caution against revising the payment formula until further research could incorporate the effects of a new payment adjustment for hospitals with high indigent care loads (the disproportionate share, or DSH, adjustment, implemented in 1985) and until the stability of their results could be examined over time. It is in that spirit that we offer the following update to their findings.

2. Data

We re-estimated hospital costs using models that are similar to HCFA’s original model and to those developed by Rogowski and Newhouse, using 2 years of data from cost reports filed for the federal fiscal years 1994 and 1995. Costs are analyzed for all short-stay facilities paid under PPS, with at least 50 Medicare discharges. Quarterly moving averages of the PPS Input Price Index were used to deflate mean operating costs per discharge. Medicare’s Case-Mix Index Files and Wage Index Files provided control variables for case severity and regional variation in factor prices. The ratio of interns and residents to beds (IRB) for each teaching hospital was derived from the indirect medical education payments reported on the settlement pages of the cost reports. A total of 4988 hospitals contributed 9627 observations to the final estimation sample, which is summarized in Table 1.

<table>
<thead>
<tr>
<th>Sample description</th>
<th>All hospitals</th>
<th>Non-teaching</th>
<th>Teaching*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations</td>
<td>9627</td>
<td>7475</td>
<td>2152</td>
</tr>
<tr>
<td>Number of unique hospitals</td>
<td>4988</td>
<td>3903</td>
<td>1145</td>
</tr>
<tr>
<td>Average real operating cost per day (1987 dollars)</td>
<td>US$3553</td>
<td>US$3163</td>
<td>US$4907</td>
</tr>
<tr>
<td>Average case mix index</td>
<td>1.2490</td>
<td>1.1849</td>
<td>1.4714</td>
</tr>
<tr>
<td>Intern and resident-to-staffed bed ratio: mean</td>
<td></td>
<td></td>
<td>0.1823</td>
</tr>
<tr>
<td>Intern and resident-to-staffed bed ratio: median</td>
<td></td>
<td></td>
<td>0.0948</td>
</tr>
<tr>
<td>Average number of staffed beds (acute care only)</td>
<td>148</td>
<td>105</td>
<td>297</td>
</tr>
<tr>
<td>Average wage index</td>
<td>0.9200</td>
<td>0.8873</td>
<td>1.0336</td>
</tr>
<tr>
<td>% Located in non-MSA</td>
<td>44%</td>
<td>55%</td>
<td>6%</td>
</tr>
<tr>
<td>% Located in small MSA (population &lt; 1 million)</td>
<td>26%</td>
<td>23%</td>
<td>37%</td>
</tr>
<tr>
<td>% Located in large MSA (population &gt; 1 million)</td>
<td>30%</td>
<td>22%</td>
<td>57%</td>
</tr>
<tr>
<td>% Receiving DSH payments</td>
<td>40%</td>
<td>33%</td>
<td>65%</td>
</tr>
<tr>
<td>Mean Medicaid utilization (% days)</td>
<td>13%</td>
<td>12%</td>
<td>17%</td>
</tr>
</tbody>
</table>

*a* Defined as those receiving Indirect Medical Education Payments during the study period.
3. Methods

3.1. Variable specification

We retained the specification of the dependent variable as the natural log of the average operating cost per case. We developed four variations on the original cost model, each derived from a different formulation of the variable measuring teaching intensity. The first model minimizes parametric assumptions by specifying teaching intensity as a set of dummy variables constructed by level of the IRB ratio, using intervals of 0.07 in the value of the ratio from 0 to 0.70, and a single group for hospitals with ratios above 0.70. This results in 11 dummy variables, with the non-teaching group serving as a reference. Our purpose in presenting these semi-parametric results is to provide a basis for comparison against which we can assess the restrictions in functional form that are imposed by each of our continuous variable models. Considering the policy application of the model, a parametric specification is preferred to a specification with indicator variables — if the parametric model fits the data well — because indicator variables would introduce notches into the payment formula.

The remaining models are derived from continuous specifications of teaching intensity including:

1. the actual value of the IRB, which ranges from zero to 1.3
2. the natural log of the IRB after adding a constant of 1.0
3. the natural log of the IRB after adding a constant of 0.001.

Because the dependent variable is logged, the second specification imposes an exponential form on the model similar to one used by HCFA in 1990 to estimate the IME adjustment for the PPS capital cost payment (Phillips, 1992). The third specification is the same as HCFA’s original measure of teaching intensity, and the fourth is similar to Rogowski and Newhouse’s recommended alternative. In keeping with Rogowski and Newhouse, we also explored generalizations of each of the continuous forms, using spline functions to identify significant thresholds for differences in the slope coefficient on teaching.

Distributions of the non-zero values of each of our teaching variables are presented in Fig. 1. Four-fifths of short stay hospitals do not participate in graduate medical education. Among those that do, the distribution of the resident-to-bed ratio is still highly skewed, with the majority of teaching hospitals showing ratios that are less than 0.10. Using a log transformation with a constant smaller than 1.0 reduces the skewness, but it also increases the influence of these smaller programs in the estimation.

3.2. Estimation

Except for alterations in the design of the teaching variable, we estimated hospital cost functions that were the same as the original functions estimated by
HCFA and then examined by Rogowski and Newhouse, except that we added a
time trend variable for the second year of the data and two indicator variables to
capture any fixed cost differentials associated with minor teaching status (0 < IRB < 0.30) or major teaching status (IRB ≥ 0.30). Neither the teaching intensity measures nor the indicator variables by teaching status estimate the economic costs of teaching per se. They serve, rather, as proxy measures for cost differentials from unobserved characteristics of care delivered in teaching settings. The empirical component of the IME adjustment can be viewed as a price premium that is predicated on the assumption that these estimates reflect legitimate product differentiation. For each specification of the teaching variable, the equation is:

\[
\ln(\text{operating cost/case}) = \alpha + \beta \text{teaching intensity} + \gamma_1 \ln(\text{CMI}) \\
+ \gamma_2 \ln(\text{wage index}) + \gamma_3 \ln(\#\text{beds}) \\
+ \gamma_4 \text{small urban status} + \gamma_5 \text{large urban status} \\
+ \gamma_6 \text{year} + \varepsilon
\]  

(1)

where \( \beta \) is the vector of predictors on the set of teaching variables appropriate to each model, including the fixed teaching effects measured by the major or minor teaching status indicators, where applicable. Weighted least squares estimation was applied using analytic weights derived from the number of discharges contributing to each hospital mean, with Huber–White robust standard errors with clustering by hospital.

Without interaction or other second-order terms in the equations, all but the first of our specifications impose an assumption that the marginal effect of teaching on the logged dependent variable is linear and constant across the estimation sample. To assess the reasonableness of this assumption, we plotted results from the models using continuous teaching variables against results computed from the dummy variable model. For formal tests of log-linearity, however, spline functions were constructed from each of the three continuous teaching variables (models 2, 3 and 4) to test for differences in the teaching slope coefficients by ranges of IRB values. Spline knots were initially set at each variable’s equivalent to IRB ratios of 0.03, 0.15, 0.30 and 0.60. Because of the relatively small sub-group size the splines for highest two groups were collapsed into one. In Rogowski and Newhouse the teaching hospitals were initially divided into six equal groups to establish spline knots. The first five were later collapsed into one group, leaving one knot at 0.25 that isolated the top 17% of teaching facilities. Our upper knot is set higher because the ratios have increased over time; 20% of teaching hospitals in our sample had ratios above 0.30.

Rogowski and Newhouse found that their formulation of the teaching variable that uses a smaller constant and allows for multiple slope coefficients reduced the model’s susceptibility to bias from omitted variables. To test whether this was true in our data, all four of our models were repeated using a specification that eliminated the variable for number of hospital beds, and another that added a Medicaid utilization variable and two sets of dichotomous variables by census region and type of ownership. (The additional independent variables were chosen
to follow Rogowski and Newhouse’s expanded model as closely as possible.) We present the resulting differences in estimated teaching effects to compare the omitted variable bias across each of the specifications.

3.3. Retransformation

To avoid confusion from the different units of measurement in teaching intensity, all results are expressed in terms of the expected teaching effect. This is defined as the percent difference in expected costs between teaching hospitals and non-teaching hospitals (assuming similarity in all covariates other than teaching) and can be expressed as:

\[
\left( \frac{E(Y)_{\text{teach}}}{E(Y)_{\text{non-teach}}} - 1 \right) = \left( \frac{e^{(X\beta)}e^{X\varepsilon}_{\text{teach}}}{e^{(X\beta)}e^{X\varepsilon}_{\text{non-teach}}} - 1 \right)
\]

(Holding other covariates constant and assuming (for now) constant variance, the numerator and denominator terms for \(e^{X\gamma}\) and \(e^{\varepsilon}\) cancel out. In all but the \(\ln(0.001 + \text{IRB})\) specification, the denominator of the right-hand side equals one, and the expected teaching effects can be measured by \(e^{(X\beta)} - 1\), where \(X\) represents the teaching variable as it is entered into the estimation of logged cost. In the case where teaching is specified as \(\ln(0.001 + \text{IRB})\), however, the denominator in the ratio of expected costs will always retain a term equal to \(e^{\ln(0.001)}\), or \((0.001)^{\beta}\).

Because Rogowski and Newhouse identified systematic differences in the distribution of the variance across teaching hospital subgroups, we could not assume that the exponentiated error terms in the above ratios would cancel out. We computed smearing estimates for the expected value of the re-transformed error computed from the mean of the exponentiated residuals, appropriate when the error is assumed to be non-normally distributed (Duan, 1983; Manning, 1998).

Let \(s\) denote smearing estimates as computed across the residuals from the appropriate group of observations. The formulas for expected teaching cost differentials in our four models can be summarized as follows:

Using dummy variables

\[
\left( 1 - \frac{s_{\text{teach}}}{s_{\text{non-teach}}} \right) - 1
\]

(3)

Using an exponential function

\[
\left( e^{(\text{IRB})\beta} e^{\mu_{\text{non}} \times \frac{s_{\text{teach}}}{s_{\text{non-teach}}}} \right) - 1
\]

(4)

Using a log-linear function, on \((1 + \text{IRB})\)

\[
\left( 1 + \text{IRB} \right) \beta e^{\mu_{\text{non}} \times \frac{s_{\text{teach}}}{s_{\text{non-teach}}}} - 1
\]

(5)
using a log-linear function, on \((0.001 + \text{IRB})\)

\[
\left[ \frac{(0.001 + \text{IRB})^{\beta_{\text{cat}}} e^{\beta_{\text{status}}}}{(0.001)^{\beta}} \times \left( \frac{s_{\text{teach}}}{s_{\text{non-teach}}} \right) \right] - 1
\]

where \(\beta_{\text{cat}}\) is the coefficient on the dummy variable and \(\beta_{\text{status}}\) is the coefficient on the major or minor teaching status indicator, as appropriate. Eq. (3) includes an additional variance correction factor for \(\beta_{\text{cat}}\) as recommended by Kennedy (1981), but where the standard error of the dummy coefficient is small, this correction has only a very slight influence on the results.

If groupwise heteroscedasticity is not present in our data, the ratio of the two smearing estimates will approximate one. If the expected teaching effect is small, however, a factor very close to 1.0 can still have a large proportional effect on the model interpretation and the resulting payment adjustments. To establish the necessity of retaining the retransformation correction factors in the computations, we therefore subjected the exponentiated residuals to further tests to evaluate whether heteroscedasticity is systematic with respect to the size of the teaching program.

4. Results

4.1. Comparing estimated teaching cost differentials across specifications

Cost differentials computed from the dummy variable model increase steadily in strength from the first through the tenth category of teaching intensity. All groups but the first test significantly different from zero at a \(p \leq 0.001\) level, and in the first category (IRB \(\leq 0.07\), mean value approximately 0.03) the parameter is still marginally significant \((p = 0.063)\). Fig. 2 plots the estimated teaching cost differentials from each of the continuous specifications against those computed from the dummy variables on categories of teaching, where bars representing the expected teaching effect by IRB category are placed approximately at the mean value of the ratio within that category. Regression results are provided in Tables 2 and 3.

There are positive but insignificant fixed teaching effects in both the exponential model and the log–log model based on \((1 + \text{IRB})\). Within the range of ratios applicable to the great majority of teaching hospitals, expected teaching effects are very similar between the exponential form and the logarithmic form that uses a base of \((1 + \text{IRB})\). The two curves do not diverge noticeably until IRB values exceed 0.75, a range which includes less than 3% of teaching institutions. Even at
a ratio of 0.91, the cost differential predicted by the exponential function is 30.9%, compared to an effect of 28.2% predicted by the model using ln(1 + IRB). Both of
Table 2  
Results from regression using categorical teaching intensity variables — parameter estimates on teaching variables only

<table>
<thead>
<tr>
<th>Dependent variable: ln(real operating cost per case)</th>
<th>Mean IRB ratio within category</th>
<th>Number observations within category</th>
<th>Robust standard error</th>
<th>Smearing estimates Within group</th>
<th>Smearing estimates Ratio to non-teach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teaching variable is 0/1, by membership in resident/bed ratio (IRB) group: Reference group is IRB = 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. IRB ratio 0–0.07</td>
<td>0.029</td>
<td>928</td>
<td>0.016</td>
<td>0.0081</td>
<td>1.052</td>
</tr>
<tr>
<td>2. IRB ratio 0.07–0.14</td>
<td>0.101</td>
<td>354</td>
<td>0.038</td>
<td>0.0118</td>
<td>1.065</td>
</tr>
<tr>
<td>3. IRB ratio 0.14–0.21</td>
<td>0.173</td>
<td>246</td>
<td>0.056</td>
<td>0.0124</td>
<td>1.039</td>
</tr>
<tr>
<td>4. IRB ratio 0.21–0.28</td>
<td>0.242</td>
<td>170</td>
<td>0.078</td>
<td>0.0174</td>
<td>1.048</td>
</tr>
<tr>
<td>5. IRB ratio 0.28–0.35</td>
<td>0.313</td>
<td>105</td>
<td>0.085</td>
<td>0.0212</td>
<td>1.057</td>
</tr>
<tr>
<td>6. IRB ratio 0.35–0.42</td>
<td>0.382</td>
<td>63</td>
<td>0.117</td>
<td>0.0205</td>
<td>1.044</td>
</tr>
<tr>
<td>7. IRB ratio 0.42–0.49</td>
<td>0.452</td>
<td>56</td>
<td>0.145</td>
<td>0.0288</td>
<td>1.060</td>
</tr>
<tr>
<td>8. IRB ratio 0.42–0.56</td>
<td>0.528</td>
<td>55</td>
<td>0.161</td>
<td>0.0303</td>
<td>1.057</td>
</tr>
<tr>
<td>9. IRB ratio 0.56–0.63</td>
<td>0.597</td>
<td>55</td>
<td>0.186</td>
<td>0.0312</td>
<td>1.053</td>
</tr>
<tr>
<td>10. IRB ratio 0.63–0.70</td>
<td>0.661</td>
<td>45</td>
<td>0.230</td>
<td>0.0212</td>
<td>1.039</td>
</tr>
<tr>
<td>11. IRB ratio 0.70–1.30</td>
<td>0.827</td>
<td>86</td>
<td>0.234</td>
<td>0.0240</td>
<td>1.030</td>
</tr>
</tbody>
</table>

(Weighted Least Squares with Huber–White correction).  
$F_{(17,4987)} = 742.94; \ p < 0.00001; \ R^2 = 0.7609$.  
(Other covariates omitted).
### Table 3
Results from regressions using alternative continuous specifications of teaching intensity variables

<table>
<thead>
<tr>
<th>Dependent variable: Inf(operating cost/case)</th>
<th>Parameter estimates by teaching specification</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IRB</td>
<td>$p &gt;</td>
<td>r</td>
<td>$</td>
<td>ln(1 + IRB)</td>
<td>$p &gt;</td>
</tr>
<tr>
<td>Teaching ratio</td>
<td>0.271</td>
<td>&lt; 0.0001</td>
<td>0.362</td>
<td>&lt; 0.0001</td>
<td>0.0196</td>
</tr>
<tr>
<td>ln(case mix)</td>
<td>0.858</td>
<td>&lt; 0.0001</td>
<td>0.859</td>
<td>&lt; 0.0001</td>
<td>0.867</td>
</tr>
<tr>
<td>ln(wage index)</td>
<td>0.674</td>
<td>&lt; 0.0001</td>
<td>0.676</td>
<td>&lt; 0.0001</td>
<td>0.681</td>
</tr>
<tr>
<td>ln(acute beds)</td>
<td>0.078</td>
<td>&lt; 0.0001</td>
<td>0.078</td>
<td>&lt; 0.0001</td>
<td>0.078</td>
</tr>
<tr>
<td>MSA &gt; 1 m population</td>
<td>0.020</td>
<td>0.051</td>
<td>0.020</td>
<td>0.048</td>
<td>0.018</td>
</tr>
<tr>
<td>MSA &lt; 1 m population</td>
<td>0.032</td>
<td>&lt; 0.0001</td>
<td>0.032</td>
<td>&lt; 0.0001</td>
<td>0.030</td>
</tr>
<tr>
<td>PPS Year 12 indicator</td>
<td>$-0.049$</td>
<td>&lt; 0.0001</td>
<td>$-0.049$</td>
<td>&lt; 0.0001</td>
<td>$-0.049$</td>
</tr>
<tr>
<td>Major teaching (IRB ≥ 0.30)</td>
<td>0.016</td>
<td>0.484</td>
<td>0.009</td>
<td>0.695</td>
<td>0.039</td>
</tr>
<tr>
<td>Minor teaching status (0 &lt; IRB &lt; 0.30)</td>
<td>0.009</td>
<td>0.260</td>
<td>0.002</td>
<td>0.771</td>
<td>$-0.064$</td>
</tr>
<tr>
<td>Constant term</td>
<td>7.608</td>
<td>&lt; 0.0001</td>
<td>7.610</td>
<td>&lt; 0.0001</td>
<td>7.743</td>
</tr>
</tbody>
</table>

| No. of observations | 9627 | 9627 | 9626 |
| No. of hospital clusters | 4988 | 4988 | 4988 |

$F$-statistic: $F_{0.4967}$ | 1347.10 | 1351.98 | 1333.65 |
$R^2$ | 0.761 | 0.761 | 0.758 |

Smearing estimates: (mean of exponentiated residuals):

- In full sample: 1.045
- In teaching only: 1.043
- In non-teaching: 1.053

(Weighted Least Squares with Huber–White adjustment; robust standard errors appear in parentheses).

These continuous specifications track closely with the results from the dummy variable model, until the highest two groups. The number of observations in these last two categories is relatively small (45 and 70, respectively) and their coefficients are well within a single standard error of each other. We could not, however, identify any individual observations that appeared to influence the
estimates for either of these two groups, and the same pattern was present in each of the two years when we examined them separately.

From the third frame in Fig. 2, it is evident that the logarithm function based on \((0.001 + \text{IRB})\) produces a substantial distortion in the teaching effect when it is restricted to a single slope coefficient for both minor and major teaching hospitals. Strong marginal effects are estimated for hospitals with very low ratios (less than 0.02) while increments above 0.10 produce little or no change in the cost differentials. Although the indicator variables for minor and major teaching status mitigate the distortion somewhat, the function does not approximate the results from the semi-parametric model.

This distortion is an artifact of the log transformation that gives additional influence to the smaller values in the distribution, and it can be corrected by relaxing the assumption of a single slope in the log form. Results from piecewise regression using spline functions are plotted in Fig. 3(a) and show that once multiple slope coefficients are allowed, substantive differences in the estimated teaching effects across the three specifications appear only at very low and very high ratios. Fig. 3(b) provides greater detail for IRB values below the first spline knot. \(F\)-tests for the equality of the spline coefficients indicate that for variables constructed from \(\ln(0.001 + \text{IRB})\), slopes are significantly different beginning with the knot at \(\text{IRB} > 0.15\) and again at \(\text{IRB} > 0.30\). There are no significant differences between slopes, however, when the spline functions are constructed from \(\text{IRB}\) without transformation, or from \(\ln(1 + \text{IRB})\).

4.2. Isolating retransformation effects

Smearing factors that are computed from the estimation sample as a whole are between 1.045 and 1.046 for all of our specifications, from which we can conclude that failure to account for error retransformation in any given prediction would result in an understatement of predicted cost by 4% to 5%. Smearing factors were also computed separately from the residuals within each teaching hospital subgroup for which the model identified different slopes and/or intercepts. Factors computed within hospital sub-groups follow similar patterns across all specifications. The ratios of smearing factors (teaching hospital subgroups divided by non-teaching hospitals) range only from 0.99 to 1.02. In comparing our ratios to those computed by Rogowski and Newhouse for hospitals grouped by low, medium and high tertiles of IRB, we find that ours are smaller. Further, we find that they are not similarly above or below 1.00; that is, the implied retransformation corrections to the payment formula are not necessarily in the same direction as those computed by Rogowski and Newhouse.

The estimated cost differentials that were plotted in Fig. 3 were simplified by assuming that the smearing estimates for the teaching and non-teaching hospitals were not statistically different (in which case, they can be assumed to have
Fig. 3. Expected teaching cost differentials computed from spline functions, across alternative continuous specifications of the teaching ratio. (a) Showing all ranges of IRB. (b) Expanded to show range below lowest spline knot.

cancelled each other out and have had no effect the calculation). When we include the actual smearing estimates in the calculations, discontinuities are introduced to the regression results. Fig. 4 illustrates this using the piecewise linear regression on \( \ln(0.001 + \text{IRB}) \), for the lower ranges of the teaching ratio where the discontinuities are most pronounced. Even though the ratios of smearing factors for the different groups of teaching hospitals are all very close to one, they do not have a trivial effect on the calculation of expected teaching cost differentials.
In the categorical model the smearing ratio from the first dummy variable group is 1.008, but including this factor in the computation increases the teaching adjustment from 1.6% to 2.4%. In the same model, the smearing ratio from the tenth group is 0.995, and including this factor lowers the adjustment from 25.8% to 25.2%. Although the proportional effects are small in the larger teaching programs, the dollar impact would still be quite large, as these hospitals have the highest IME adjustment rates and the highest volume of Medicare cases.

Given the magnitude of the impact of these corrections on both distribution and total level of IME payments, a formal test of the assumption of heteroscedasticity is in order. We repeated one of Rogowski and Newhouse’s tests for non-constant variance by regressing the exponentiated residuals on a constant plus sets of dummy variables reflecting hospital sub-groups as they were defined in each of our specifications. In each case, the coefficients on the dummy variables tested individually and jointly non-significant. The $F$-statistic for the joint null hypothesis on the dummy variable model was $F_{(1,9615)} = 0.26$ ($p = 0.99$); on splines from IRB, $F_{(4,9621)} = 0.69$ ($p = 0.63$); on splines from $\ln(1 + \text{IRB})$, $F_{(4,9621)} = 0.68$ ($p = 0.64$); on splines from $\ln(0.001 + \text{IRB})$, $F_{(4,9621)} = 0.53$ ($p = 0.75$). In contrast, from the same test on a model derived from splines on $\ln(0.0001 + \text{IRB})$, Rogowski and Newhouse obtained $F_{(3,5064)} = 5.49$, $p < 0.01$. Thus, groupwise heteroscedasticity with respect to teaching activities, which was identified by this and other tests in the Rogowski–Newhouse study, does not appear to be present in the 1994–1995 data.
4.3. Sensitivity to omitted variables and outlier cases

We are interested in whether any of our specifications appears to be less sensitive than the others to added variables or to extreme values or influential observations. All of the specifications of the hospital cost function suffer from omitted variable bias. For example, in the model using splines derived from \( \ln(0.001 + \text{IRB}) \), at an IRB of 0.25, adding controls for Medicaid utilization, census region and ownership decreases the estimated teaching cost differentials by 3.1 percentage points (from 7.6% to 4.5%). Similar increments are computed, however, when we use the other two spline functions. Excluding beds from the equation has the opposite effect on the teaching coefficients, but again, effect is similar in each of the three splines. As is evident from the four frames in Fig. 5,
the spline versions of both log–log models are sensitive to the exclusion or inclusion of key hospital characteristics. There does appear to be some difference in sensitivity to added variables between the single slope and the multiple slope formulations on the ln(1 + IRB), but it is not consistent; at lower IRB levels the single-slope estimates are less sensitive, while at higher levels they are more sensitive.

Our results contrast sharply to what Rogowski and Newhouse found in similar variations. They estimated, for example, that excluding beds from the original single slope model using ln(1 + IRB) increased the cost differentials by 3.5 percentage points. Yet when they excluded beds from a model using their final re-formulation of the teaching variable, estimated teaching differentials increased by only 0.8 percentage points.

All of our models are similar in predictive ability, with $R^2$ between 0.76 and 0.77, with or without the added variables. This is not surprising, as over 70% of variation in logged operating cost per case can be explained by case mix and wage index values alone. Although this is an improvement over the explained variance in the estimations on 1984 data (from 0.65 to 0.68) the most likely explanation for this improvement is reduced measurement error in the teaching, case-mix and wage index measures over this period. Refining the model with additional explanatory variables alters the parameter estimates on the original covariates, but it has little effect on predicted costs. We do not, however, address the conceptual or policy motivations for HCFA’s decision to restrict the model’s covariates; see Sheingold (1990).

Examination of the residuals revealed few differences among the three continuous teaching variable models once multiple slope coefficients were introduced. In all cases outlier observations were predominantly non-teaching hospitals—either very small institutions or high-cost specialty facilities, with extreme values in the dependent variable. Because cost per case has been logged, their impact on the model is already minimized and predicted results are not sensitive to their inclusion in the estimation sample. High leverage observations tended to be more evenly distributed across the sample by teaching status. Considering the immediate policy applications of this model, we are more interested in identifying observations with undue influence on the teaching coefficient than on predicted values. We examined sensitivity to extreme values in the explanatory variable of interest by excluding the few observations at the far right of the IRB distribution (above 1.0). This change had only a slight effect on the teaching coefficient in the exponential model and negligible effects on the logged specifications. We also constructed a wage and case-mix adjusted cost per case variable (thus standardizing for the model’s two most powerful covariates), from which we identified observations that were greater than three standard deviations from the geometric mean. In estimations that excluded these potentially influential observations, the coefficients were still very similar to those generated by the full sample, for the teaching as well as other independent variables, across all of our specifications.
5. Discussion

The differences that we have identified between our findings and those of Rogowski and Newhouse may be due to secular changes in the hospital industry over the period from 1984 to 1995, some of which is evident from the parameter estimates generated on the model’s control variables. These are different in magnitude, significance, and even direction from effects measured on similar variables from other studies using earlier data, including those by Rogowski and Newhouse. In both the original and expanded models, for example, the differentials between urban and rural areas have declined sharply and the magnitude of the teaching coefficient has declined. In the expanded model (regression results available from the authors on request), the coefficients on public and private ownership reversed sign.

While our findings on all continuous teaching variable specifications other than those with the smaller constant indicate that the log-linear effect in 1994–1995 is constant throughout the sample, and that there are no significant fixed differentials for major or minor teaching institutions, this was not true in earlier critiques of the cost function (Welch, 1987; Thorpe, 1988). More recently, a panel study conducted by the authors with data from 1989 through 1995 identified significant positive interaction effects between teaching intensity and both teaching status and DSH eligibility, when measured across the seven-year period (Dalton et al., 2000). The strength of the interactions declined over time, however, and they are no longer significant by the last years of data.

In the Rogowski and Newhouse study it is not possible to separate the effect of allowing multiple slope coefficients from the effects of altering the log transformation, since multiple slopes were not tested under the other formulations of the teaching variable. Thus, the improved stability that they noted in their model may be attributable to the generalization of the log-linear form through the spline functions during a period when the slope coefficients were truly different—rather than to the use of the smaller constant.

Several conclusions can be drawn from our findings that could help guide Medicare’s re-estimations for payment differentials to teaching hospitals. With respect to the issue of specifying an appropriate continuous measure of teaching intensity, we find no compelling econometric reason to alter HCFA’s original formulation of the variable as the natural log of one plus the teaching ratio. From the 1994–1995 data, there is no evidence that adding a smaller constant to the ratio introduces less bias to the estimation, or that it makes the estimation of teaching effects less sensitive to the influence of other covariates. On the contrary, we find that the smaller constant introduces distortion to the model that can only be eliminated by relaxing the assumption of constant elasticity, forcing an otherwise unnecessary level of complication on the model. It is possible that when Rogowski and Newhouse conducted their study using 1984 hospital costs, the coefficient on the teaching variable was truly different by levels of teaching.
intensity. In that case their formulation of the teaching variable using splines derived from \( \ln(0.0001 + IRB) \) would have corrected an underlying error in the functional form of the HCFA model, but for reasons relating to the spline formulation rather than the use of a smaller constant.

From the results on the model using dummy variables, we do not identify a threshold below which no significant teaching effect exists. Cost differentials for hospitals with ratios below 0.07 are small but marginally significant. Models using a continuous teaching measure with multiple slopes consistently produce non-significant coefficients at very low IRB values. There may, however, be insufficient variation in the lower ranges of the ratio to identify the effect in a continuous measure, despite the large number of observations in these groups. From the plotted results of the dummy variable model we can also see that an assumption of decreasing marginal effect in the higher ranges of the teaching ratio appears more likely to fit the data than one of increasing marginal effect, even though the small number of observations with ratio values above 0.70 prevents us from rejecting the exponential model on the basis of statistical tests. HCFA currently uses an exponential model to compute IME adjustments to the capital portion of the PPS payment, based on the ratio of residents to average daily census. If an exponential model were used to compute adjustments for both the operating and capital components of the per-case payments, it would produce substantially higher payments for a small group of hospitals with very large teaching programs. We do not, however, find any empirical evidence that the exponential model represents an improvement over the original logarithmic one. Results from the semi-parametric model might even suggest an adjustment formula that is derived from either the IRB or \( \ln(1 + IRB) \), but is capped at some maximum value around 0.70 or 0.80.

Finally, our findings indicate that including separate smearing estimates in the computation of expected teaching cost differentials is unnecessary. Although variance correction factors are appropriate to retransformation of logged model results, unless there are systematic differences in the distribution of the error term across groups of teaching hospitals, the factors will cancel each other out in the IME payment formula. We find no evidence of such systematic differences in our data. Rogowski and Newhouse correctly identified the retransformation error that was present in HCFA’s interpretation of the teaching coefficient, in light of the significant groupwise heteroscedasticity that they identified in 1984 data. Therefore, because this issue has such a strong potential influence on estimated teaching cost differentials, any future re-estimations of indirect medical education costs should continue to test for this type of non-constant variance.

Acknowledgements

Corresponding author was supported through funding from the Agency for Health Care Policy and Research, National Research Service Award post-doctoral fellowship.
References


