Identifying and Accommodating Statistical Outliers When Setting Prospective Payment Rates for Inpatient Rehabilitation Facilities

Susan M. Paddock, Barbara O. Wynn, Grace M. Carter, and Melinda Beeuwkes Buntin

Objective. To demonstrate how a Bayesian outlier accommodation model identifies and accommodates statistical outlier hospitals when developing facility payment adjustments for Medicare's prospective payment system for inpatient rehabilitation care.

Data Sources/Study Setting. Administrative data on costs and facility characteristics of inpatient rehabilitation facilities (IRFs) for calendar years 1998 and 1999.

Study Design. Compare standard linear regression and the Bayesian outlier accommodation model for developing facility payment adjusters for a prospective payment system.

Data Collection. Variables describing facility average cost per case and facility characteristics were derived from several administrative data sources.

Principal Findings. Evidence was found of non-normality of regression errors in the data used to develop facility payment adjustments for the inpatient rehabilitation facilities prospective payment system (IRF PPS). The Bayesian outlier accommodation model is shown to be appropriate for these data, but the model is largely consistent with the standard linear regression used in the development of the IRF PPS payment adjusters.

Conclusions. The Bayesian outlier accommodation model is more robust to statistical outlier IRFs than standard linear regression for developing facility payment adjustments. It also allows for easy interpretation of model parameters, making it a viable policy alternative to standard regression in setting payment rates.

Key Words. Bayesian statistics, inpatient rehabilitation care, Medicare, outliers, prospective payment system

As part of the Balanced Budget Act of 1997, the U.S. Congress mandated that the Centers for Medicare and Medicaid Services (CMS) implement a prospective payment system (PPS) for hospital inpatient rehabilitation facilities (IRF) who care for Medicare beneficiaries. In 1999, Medicare's payments for inpatient rehabilitation care were \$4.2 billion. This new IRF PPS was implemented on January 1, 2002. Under the new IRF PPS, inpatient rehabilitation

facilities are compensated for providing inpatient rehabilitation care based on a predetermined amount per case according to the patient's impairment, age, level of function, and comorbid conditions. As is the case for hospitals under the acute care PPS, some cost differences among IRFs are due to factors other than patient characteristics, such as wage differences across geographic areas, and the fact that payment under the new IRF PPS is also a function of facility-specific factors.

A standard method for adjusting payment rates for facility cost differences in Medicare prospective payment systems is to use standard linear regression to model the average case-mix-adjusted cost per case at a facility as a function of facility characteristics that are related to cost. Functions of the resulting coefficients are used as payment adjustors (Lave 1984; Sheingold 1990). Several features of these "payment regressions" have been discussed in the literature. For example, Sheingold (1990) focuses on the use of linear regression in developing payment adjustments for the Medicare acute hospital PPS, stating that the variables included in the payment regression must not only be related to cost, but must also be beyond the control of the facilities so as not to introduce inappropriate incentives into the payment system. Other authors have focused on the impact of various transformations of predictor variables in the payment model; Rogowski and Newhouse (1992) and Dalton and Norton (2000) discuss the impact of the specification of the indirect teaching variable on Medicare's acute PPS.

A key assumption of standard linear regression is that regression error terms are normally distributed; the linear regression framework might be inappropriate for payment adjustments if this assumption is violated. This article focuses on the validity of the assumption of normally distributed errors using the derivation of facility payment adjustments for the IRF PPS as an example. The article is structured as follows. First, background is provided on the IRF PPS and the role of facility payment adjustments within this payment system. An overview of the standard method for deriving facility payment adjustments is given, followed by an illustration using the data that were used to derive the facility payment adjustments for the IRF PPS. We show that the assumption of normally distributed regression errors is questionable and

This work was funded by the Centers for Medicare and Medicaid Services (Contract #500-95-0056).

Address correspondence to Susan M. Paddock, RAND, 1700 Main St., Santa Monica, CA 90407-2138. Barbara O. Wynn and Melinda Beeuwkes Buntin are with RAND, Arlington, VA. Grace M. Carter is with RAND, Santa Monica, CA.

discuss alternative statistical approaches to address this issue. In particular, we propose using a Bayesian outlier accommodation model to develop payment adjustments. The Bayesian outlier accommodation model relaxes the assumption of normally distributed errors, thus accommodating potentially unduly influential IRFs in the analysis. We show that it is structurally similar to the standard linear regression model, thus readily fitting into the existing paradigm for developing facility payment adjustments. Our discussion focuses on the IRF PPS, but the issues raised here also apply to the linear regression models used to set facility payment adjustments in other prospective payment systems.

Note that the term “statistical outlier” is used throughout this paper to denote IRFs that are identified as being highly unusual relative to other IRFs that are part of the statistical analysis. The statistical outlier IRFs that are discussed here have nothing to do with Medicare’s outlier payment policies, which pertain not to facilities but rather to cases that have unusually high costs as determined by criteria that are specific to Medicare’s various payment systems.

DESCRIPTION OF THE IRF PPS

The unit of payment in the IRF PPS is a Medicare covered hospital stay, beginning with an admission to the rehabilitation hospital or a rehabilitation unit of an acute care hospital and ending with discharge from that facility. Each case is classified into one of 380 case-mix groups. The case-mix groups (CMGs) are based on cause of impairment, age, functional status as measured by items based on the FIM™ instrument (Uniform Data System for Medical Rehabilitation 1997), and comorbid condition, with additional groups constructed for deaths and atypically short-stay cases. A weight is associated with each CMG that reflects the relative costliness of a case within that CMG. The payment for a rehabilitation stay under the IRF PPS is a product of three factors: a national baseline payment; the weight assigned to the patient’s case-mix group that adjusts the national baseline payment for the relative costliness of the patient; and a facility adjustment to compensate IRFs for factors associated with increased costs that are beyond the control of the IRFs. In addition, the payment is adjusted for outlier cases and short stay transfer cases. The development of the standard baseline payment, case weights, and rules for payment of unusual cases are further detailed in Carter et al. (2002).

DATA FOR FACILITY ADJUSTMENTS FOR THE IRF PPS

Carter et al. (2002) used data on costs and characteristics for 714 IRFs from calendar years 1998 and 1999 to develop the IRF PPS facility payment adjustments. The cost for each discharge was estimated as the sum of (1) the estimated costs incurred by the patient in each ancillary department as determined by applying a departmental cost-to-charge ratio from the Healthcare Cost Report Information System (HCRIS) to the patient's charges in the department as reported in the Medicare Provider Analysis and Review (MEDPAR) claims file, and (2) the average per diem costs for "room and board" multiplied by the patient's length of stay (Newhouse, Cretin, and Witsberger 1989). The average cost per case for an IRF was computed by summing the costs for all cases in the sample and dividing by the number of equivalent full cases, where an equivalent full case is defined in a manner consistent with the payment policy of the IRF PPS (wherein a short stay transfer counts as only a fraction of a case). The IRF average cost per case was pooled across 1998 and 1999.

Facility characteristics for the IRFs are defined in Table 1 along with their mean values and standard deviations, which are weighted by the number of equivalent cases per facility. The case-mix index, which is a measure of the average resource use and impairment of cases at an IRF, was computed as the

Table 1: IRF Characteristics List for the Regressions

<i>Variable Name</i>	<i>Definition</i>	<i>Mean (Std. Dev.)</i>
Log(CMI)	log(case mix index [CMI])	- 0.011 (0.137)
Wage index	log(0.295+0.705*CMS wage index measure)	- 0.025 (0.099)
Low income	log(1+measure of low-income patients)	0.112 (0.072)
Low-income imputation	imputation indicator for log1_dsh (yes = 1)	0.233 (0.423)
Indirect teaching	log(1+interns to residents ratio)	0.013 (0.042)
Freestanding IRF	freestanding rehabilitation facility (yes = 1)	0.416 (0.493)
Small IRF	Freestanding IRF with average daily census (adc) <25 or a unit with adc < 10 (yes = 1)	0.112 (0.315)
Large IRF	Freestanding IRF with adc ≥ 50 or a unit with adc ≥ 25 (yes = 1)	0.460 (0.499)
Old certification date	certification date before 01/01/1985 (yes = 1)	0.343 (0.475)
New certification date	certification date after 12/31/1990 (yes = 1)	0.292 (0.455)
Large urban location	large urban (yes = 1)	0.460 (0.499)
Rural location	rural (yes = 1)	0.070 (0.255)
Proprietary IRF	proprietary control of facility (yes = 1)	0.364 (0.482)

facility average of the CMG weights for its cases. The other variables in Table 1 were derived from HCRIS, Online Survey, Certification and Reporting System (OSCAR), or were provided directly by the Centers for Medicare and Medicaid Services. The wage index variable was derived from a wage index measure provided by CMS, which is based on fiscal year 1997 cost report hospital wage data for all hospitals. The wage index variable is computed to reflect the CMS actuarial assumption that labor-related costs comprise 70.5 percent of all hospital costs; that is, the wage index variable used in the regressions is the logarithm of $(0.705 * \text{CMS wage index measure} + 0.295)$. The low-income percentage variable is the percentage of Medicare patients who are entitled to Supplemental Security Insurance plus the percentage of all inpatients eligible for Medicaid (but not Medicare); this measure is consistent with the low-income measure used in the PPS for acute care hospitals. This low-income variable was imputed for some of the IRFs, and an indicator of whether this variable was imputed is included in the regression models. The indirect teaching measure is computed as the ratio of full-time equivalent (FTE) residents to average daily census. The certification date variables indicate when an IRF began accepting Medicare patients. Newly established IRFs could have higher costs per case because they had several years to establish their base costs before a rate-of-increase cost limit was established under the former payment system. An indicator variable as to whether an IRF is a freestanding hospital or a unit of a general acute care hospital is also included in the regression models. Variables indicating location in a large urban area (defined as Metropolitan Statistical Areas [MSAs] with 1 million population or more) or in a rural (non-MSA) area are also included.

FULLY SPECIFIED AND PAYMENT REGRESSIONS

Sheingold (1990) differentiates between *fully specified* and *payment* regressions. The goal of the fully specified regression is to gain an understanding of the relationship between average cost per case and facility characteristics, regardless of whether the facility characteristics in the model are candidate payment adjusters. The payment regression model is used to develop the adjusters actually used in the payment system. The goal of the fully specified regression is strictly statistical; factors in the fully specified model deemed to be related to cost but not appropriate for direct compensation are omitted from the payment regression model and are compensated only to the extent that they are correlated with those effects that are paid for under the payment

system (Sheingold 1990). The first step in developing payment adjustors for the IRF PPS is to fit the fully specified regression of the logarithm of the average cost per case at IRF i on the facility characteristics that are listed in Table 1 to understand which variables were significantly related to $\log(\text{cost})$. The regression errors, ε_i for the IRFs are assumed to be independent and normally distributed with means 0 and variances σ^2/n_i :

$$\varepsilon_i \sim N(0, \sigma^2/n_i) \tag{1}$$

where n_i is the number of equivalent cases at IRF i ; a weighted regression is performed to reflect the per discharge payment rule of the IRF PPS (Sheingold 1990). Table 2(a), which comes from Model 2 (Table 7.8 of Carter et al. 2002) shows the regression coefficients for this regression model. Case-mix index, wage index, and low-income percentage are positively and significantly related to cost. Inpatient rehabilitation facilities that were freestanding, small, new (i.e., certification date after 1990), and located in rural areas have significantly higher costs than others, while large urban IRFs tend to have lower costs.

Table 2: Fully Specified Regression: Log Average Cost per Case Regressed on Facility Characteristics Using (a) Standard Linear Regression and (b) Bayesian Outlier Accommodation Model

Variable	(a)			(b)		
	Mean	95% Confidence Interval		Mean	95% Posterior Probability Interval	
		2.5%	97.5%		2.5%	97.5%
Log(CMI)	1.146	1.020	1.273	1.156	1.035	1.278
Wage index	1.252	1.050	1.455	1.229	1.010	1.445
Low income	0.359	0.109	0.610	0.294	0.050	0.539
Low-income imputation	0.029	− 0.010	0.069	0.029	− 0.010	0.067
Indirect teaching	0.342	− 0.064	0.748	0.438	0.052	0.820
Freestanding IRF	0.053	0.012	0.094	0.056	0.013	0.099
Small IRF	0.081	0.026	0.135	0.080	0.028	0.132
Large IRF	− 0.029	− 0.065	0.007	− 0.032	− 0.070	0.006
Old certification date	− 0.018	− 0.058	0.022	− 0.010	− 0.051	0.030
New certification date	0.048	0.008	0.088	0.058	0.017	0.098
Large urban location	− 0.050	− 0.088	− 0.012	− 0.043	− 0.082	− 0.004
Rural location	0.140	0.072	0.209	0.138	0.070	0.205
Proprietary IRF	0.029	− 0.013	0.070	0.029	− 0.014	0.072
Intercept	9.323	9.274	9.372	9.319	9.271	9.366
R ²		0.465			0.532	

Table 3: Payment Regression: CMI- and Wage Index-Adjusted Cost per Case Regressed on Payment Variables Using (a) Standard Linear Regression and (b) Bayesian Outlier Accommodation Model

Variable	(a)			(b)		
	Mean	95% Confidence Interval		Mean	95% Posterior Probability Interval	
		2.5%	97.5%		2.5%	97.5%
Low income	0.484	0.255	0.713	0.424	0.199	0.652
Low-income imputation	0.022	-0.017	0.061	0.027	-0.012	0.066
Rural IRF	0.175	0.110	0.240	0.165	0.101	0.229
Intercept	9.317	9.284	9.350	9.325	9.292	9.358
R^2		0.060			0.065	

Given the results of Table 2(a) and discussion with CMS about IRF characteristics that CMS considered appropriate for payment under the IRF PPS, a final payment regression model was derived. The logarithm of the average cost per case, standardized for IRF case-mix index and wage index (Sheingold 1990), was regressed on those terms displayed in Table 3(a) (Model 10 of Table 7.12 of Carter et al. 2002). The payment adjustment for the wage index measure is based on the labor-related share of costs and thus not derived from the regressions. Based on the regression coefficients, the other payment adjustments are 19.1 percent for rural facilities and $(1 + \text{low-income percentage})^{0.484}$ for serving low-income patients. These adjustments are derived from the regression coefficients of Table 3(a) as follows. The rural coefficient in the regression, 0.175, is exponentiated, since the outcome variable is the logarithm of the average cost per case, to obtain the rural adjustment: $\exp(0.175) = 0.191$, or 19.1 percent. Similarly, the regression coefficient for $\log(1 + \text{low-income percentage})$ is 0.484; on the nonlogged scale, the adjustment becomes $\exp(0.484 \cdot \log[1 + \text{low income percentage}]) = (1 + \text{low income percentage})^{0.484}$. Status as a freestanding rehabilitation facility, small facility size, and recent certification date are significantly related to increased cost—Table 2(a)—but are excluded from the payment regression since CMS does not have a policy reason to directly compensate IRFs for these characteristics. However, the factors that are excluded from the payment regression may be indirectly compensated for to the extent that they are correlated with those factors that are explicitly included in the payment regression (Lave 1984; Sheingold 1990).

APPROPRIATENESS OF THE NORMAL ERRORS ASSUMPTION

While the goals of the payment regression are more policy-oriented than statistical, the fully specified regression should be evaluated from a statistical viewpoint. The variables that are significant predictors of payment and thus potential payment variables are identified with the fully specified regression model.

A correct interpretation of the fully specified regression relies on basic linear regression assumptions to hold, namely that the error terms are normally distributed. Figure 1(a) is a standard regression diagnostic plot that indicates whether the errors are normally distributed. If the errors are normally distributed, then the observations should fall along the solid diagonal line in that graph. However, Figure 1(a) indicates that the normal errors assumption is questionable, given the number of observations that fall below the line on the left-hand side and fall above it on the right-hand side. This pattern indicates that an error distribution with heavier tails might be more appropriate given the data (Weisberg 1985).

When faced with the presence of statistical outlier IRFs in the data, a few options are typically employed in practice:

1. Fit the model using standard linear regression, with all observations. Potential problems with this approach are that statistical outliers could unduly influence regression coefficients and thus payment adjustments. The statistical outlier IRFs could be unusual enough that they should be excluded from the payment regression for policy reasons; for example, if the statistical outlier IRFs were not subject to the same constraints or rules as the other IRFs. For instance, IRFs that were in their phase-in periods under the former payment system during 1998 and 1999 and were thus establishing their costs could possibly be different from IRFs already subject to the payment system.
2. Delete statistical outlier IRFs from the data and then derive the facility payment adjustments. This would be reasonable if it were clear that the statistical outlier IRFs do not represent the universe of facilities on which standard IRF PPS payments should be calculated, perhaps because of the phase-in period or because of reporting errors. However, deleting certain IRFs from the calculation of the facility payment adjustment without a good reason could be criticized as subjective. If there is not a clear reason for deleting them from the data, then deleting them risks losing potentially valuable information

Figure 1: Linear Regression Diagnostic Plots for the Fully Specified Linear Regression (a) and for Errors following T-Distributions with $\nu = 50, 30, 20, 10$ and 5 Degrees of Freedom (b through f)

about the relationship between cost and IRF characteristics. Deleting IRFs from the analysis relies strongly on the normal errors assumption of linear regression; it may be the case that this assumption is unwarranted and an alternative model would be more appropriate.

3. Fit a statistical model to the data that is sufficiently robust to statistical outliers such that coefficient estimates are relatively insensitive to their presence in the data. Health policy practitioners find linear regression to be valuable because of its overall interpretability and ease of use (Sheingold 1990), so ideally an alternative to standard linear regression should be readily interpretable as well. Robust regression provides such an alternative. Robust regression procedures are available in some commonly used software packages, such as *Stata*'s "rreg" procedure (Stata Corporation 2003; Hamilton 1998). The rreg procedure first deletes gross outliers from the data and then computes observation-specific weights for the contribution of each observation to the final regression; these weights take on values between 0 and 1, such that a case with weight 1 will be fully counted while a case with a weight of 0 will be deleted from the analysis. The user specifies tuning constants prior to the analysis that determine the degree to which each observation will affect the regression coefficients in the final model. Because rreg does not allow for the facility weights to be properly included in the model to reflect the per-discharge payment policy of the IRF PPS—that is, the n_i from Equation (1)—we do not further consider it as an option for developing facility payment adjustments. We fit a Bayesian outlier accommodation model to the data, therefore, which is robust to the presence of outliers while allowing for the facility weights to be included in the analysis.

BAYESIAN OUTLIER ACCOMMODATION MODEL

We explore replacing the standard linear regression model with a Bayesian outlier accommodation model for developing payment adjustments. The Bayesian outlier accommodation model downweights the influence of statistical outlier IRFs relative to standard linear regression and allows for IRFs with unusually high variance in costs to remain in the data without unduly influencing the regression coefficients. Structurally, the Bayesian outlier accommodation model is very similar to the standard linear regression model, with the key difference being that the error term for IRF i is normally dis-

tributed with variance $\sigma^2/(\lambda_i n_i)$, rather than σ^2/n_i as in standard weighted linear regression. The λ_i term for IRF i indicates the degree to which IRF i is unduly influential in the standard linear regression analysis and the extent to which its contribution to the estimation of the regression parameters should be downweighted relative to the standard normal linear regression. In the standard linear regression, λ_i is equal to 1 for all IRFs. If IRF i is unduly influential, then λ_i is less than 1 in the Bayesian outlier accommodation model, thereby effectively reducing the IRF weight from n_i to $\lambda_i n_i$. The λ_i terms are unknown and, along with the regression coefficients, must be estimated from the data. In contrast to standard robust regression such as that offered in *Stata's* `reg`, the outlier accommodation is explicitly included in the model formulation by entering these λ_i terms into the model, as opposed to using user-specified tuning constants. Since each λ_i is always greater than 0, no observation is ever completely deleted from the analysis.

The specification of the Bayesian outlier accommodation model is completed by specifying a distribution for the λ_i . The λ_i terms are modeled as following a Gamma distribution with parameters $(v/2, v/2)$ with mean 1. The complete model specification implies that $\log(\text{cost})$ follows a t -distribution with v degrees of freedom (Verdinelli and Wasserman 1991; Gelman et al. 1996). The t -distribution is a symmetric distribution, just like the normal distribution, but it has heavier tails and allows for the existence of more unusual, higher variance data points than does the normal distribution. The degrees of freedom parameter, v , indicates how heavy tailed the distribution is relative to the normal; Figures 1(b)–(f) show the normal quantile plot for errors generated from t -distributions with 50, 30, 20, 10, and 5 degrees of freedom, denoted by “ $t(50)$ ” through “ $t(5)$ ” on Figure 1. The larger the degrees of freedom, the closer the t -distribution is to the normal distribution. In practice, t -distributions with degrees of freedom of 30 or greater are considered to be approximately normal. The t -distribution is a good choice for these data; Figure 1(a) shows the regression diagnostic plot for the fully specified standard linear regression of Table 2(a), and the figure indicates that not only are there IRFs that appear to be statistical outliers, but the assumption of normally distributed errors is questionable, as Figure 1(a) more closely resembles Figures 1(e)–(f) than Figures 1(b)–(c).

BAYESIAN ANALYSIS, COMPUTATION, AND PRIOR DISTRIBUTIONS

We reanalyze the data used to derive the facility payment adjustments for the IRF PPS using the Bayesian outlier accommodation model. Not only do we

estimate regression coefficients using this model, but we also estimate how non-normal the error distribution is, given the data. To do this, we estimate the degrees of freedom of the t -distribution, v , as part of the analysis.

Several introductory articles on Bayesian methods have appeared in the health literature (Spiegelhalter et al. 1999; Harrell and Shih 2001). Below we provide a basic description of the Bayesian approach but readers who want more information may wish to consult those articles. The Bayesian approach centers on Bayes's theorem: $p(\theta|y) \propto f(y|\theta)p(\theta)$. The first step of the Bayesian analysis is the same as that of the standard, non-Bayesian analysis: one specifies a likelihood function for the data, which is $f(y|\theta)$ in Bayes's theorem, where θ represents the model parameters (i.e., the regression coefficients, the variance, σ^2 , the degrees of freedom parameter, v , and λ). The Bayesian approach treats the parameters of the model as random variables and requires that *prior distributions* be specified for them; these prior distributions are denoted by $p(\theta)$ in Bayes's theorem.¹ The prior distribution for a given parameter quantifies the beliefs of the analyst about that parameter prior to the analysis.² The Bayesian analysis proceeds by multiplying the prior distribution and the likelihood function together to obtain the *posterior distribution* of the model parameters, $p(\theta|y)$. All Bayesian statistical inference is based on the posterior distributions of the model parameters.

The prior distributions chosen for this analysis are very diffuse with high variances to reflect our desire to express ignorance prior to the analysis about the possible values for the regression parameters. The prior distribution for each regression coefficient follows a normal distribution with mean 0 and a standard deviation of 100. This prior choice reflects that we have little prior knowledge about the values of the regression coefficients. Considering that the largest standard error for the regression coefficients in Table 2(a) is 0.21, this prior distribution is indeed very vague. The inverse of the variance, σ^{-2} , follows a Gamma distribution with parameters (0.01, 0.01). These prior distribution choices have historically been standard choices in Bayesian linear regression (DeGroot 1970), though both choices are sufficiently noninformative enough that the posterior estimates are insensitive to them in this analysis.

The model is fit using three specifications to assess the sensitivity of results to the choice of the prior distribution for v . The first specification places equal prior probability on values of v that support the assumption of approximately normal errors (of standard linear regression) versus smaller values of v that suggest the use of a heavier tailed t -distribution; the second prior places probability 0.01 on v being greater than or equal to 30; and the third places

probability 0.2 on such an event a priori. The three prior specifications for v that were examined are:

- (1) $\gamma = 1/v$ and a Beta prior is placed on γ such that the probability of v being greater than 30 is 0.5;
- (2) $\gamma = 1/v$ and a Beta prior is placed on γ such that the probability of v being greater than 30 is 0.01 (Verdinelli and Wasserman 1991);
- (3) v can take values 2, 4, 6, 8, 10, 12, 15, 20, 30, or 50 with probability $1/10$ for each possible value (Spiegelhalter et al. 1995).

We found that the posterior distributions were insensitive to these choices of priors and so only present the results based on specification (1).

The Bayesian outlier accommodation model is easy to fit in a freely available software package called *WinBUGS* Version 1.3 (Spiegelhalter, Thomas, and Best 2000). It is downloadable from <http://www.mrc-bsu.cam.ac.uk/bugs/>. A tutorial article on *WinBUGS* software is available from Fryback, Stout, and Rosenberg (2001). *WinBUGS* is especially useful for Bayesian analyses for which the posterior distributions of interest cannot be analytically derived. Within *WinBUGS*, one specifies the likelihood function and prior distributions and then *WinBUGS* derives the posterior distributions of interest by using a numerical integration technique called Markov Chain Monte Carlo (MCMC) simulation. The MCMC simulation is iterative; once starting values are specified for model parameters, MCMC simulation is used to repeatedly draw random samples from the posterior distributions of interest. Posterior means and standard deviations of model parameters can be computed directly from these MCMC samples. *WinBUGS* has several diagnostic tests housed within it to help the user assess whether the MCMC simulation algorithm has converged. The *WinBUGS* code used for the analysis can be requested from the corresponding author of this paper.

RESULTS: FULLY SPECIFIED BAYESIAN OUTLIER ACCOMMODATION MODEL

The Bayesian outlier accommodation model is fit using the same dataset used to produce Tables 2(a) and 3(a), which appear in Carter et al. (2002). Table 2(b) shows the 95 percent posterior probability intervals for the regression parameters. The 95 percent posterior probability interval is the Bayesian analogue to the 95 percent confidence interval, though note that their interpretations are

different: The Bayesian 95 percent posterior probability interval for the freestanding regression coefficient can be interpreted as the probability that the freestanding coefficient is in (0.013, 0.099) is 0.95, whereas the analogous 95 percent confidence interval is interpreted as the probability that the interval, (0.012, 0.094), contains the true, fixed value of the freestanding coefficient is 0.95. The regression coefficients of Tables 2(a) and 2(b) largely agree. The posterior mean of v is 7.9, with a 95 percent posterior probability interval of (4.7, 14.1), which provides evidence that the correct error distribution is indeed fatter-tailed than a normal (recall from Figure 1 that values of v greater than 30 indicate approximate normality of the regression errors). Hence, the two columns labeled "mean" in Table 2 differ because the assumed error distributions differ, which determines how the extreme observations affect regression parameter estimates. Though the posterior mean for low income measure is lower than the estimate in Table 2(a) (0.294 vs. 0.359), this variable is significantly related to cost in both models, as the 95 percent confidence interval of Table 2(a) and the 95 percent posterior probability interval of Table 2(b) fall above 0. One disagreement between Table 2(a) and Table 2(b) is that the indirect teaching variable is significantly related to $\log(\text{cost})$ in Table 2(b) but is not significant in Table 2(a); the posterior mean of the indirect teaching coefficient is 0.438 in Table 2(b), with 95 percent posterior probability interval of (0.052, 0.820), whereas in Table 2(a) the estimate is 0.342 with 95 percent confidence interval of $(-0.064, 0.748)$. This is particularly interesting because characteristics that were significantly related to cost after adjusting for other facility characteristics were more carefully considered for inclusion in the payment regression model; teaching is compensated under the acute care PPS; and there were issues regarding the availability of a reliable resident count measure. In particular, full-time equivalent counts for number of residents in freestanding IRFs and in rehabilitation units are not provided on the cost report. To be comparable with the number of residents reported for a rehabilitation unit, the number of residents assigned to the routine area of a freestanding hospital was estimated by multiplying its total resident count by the ratio of resident salaries apportioned to routine areas to the total resident salaries for the facility (Carter et al. 2002).

IDENTIFICATION OF EXTREME OBSERVATIONS

The λ_i parameters indicate which IRFs are statistical outliers and will therefore be downweighted in the analysis. An analyst might be interested to see if IRFs

that are flagged as statistical outliers by having lower values of λ_i are associated with certain features of IRFs. For example, 23 IRFs had low posterior probabilities of λ_i being equal to or greater than 1. In particular, the posterior probability that λ_i was equal to or greater than 1 was less than 0.10 for these IRFs; in other words, each IRF had a probability greater than 0.90 of being downweighted. In this analysis, this characterization of outlier IRFs coincides to selecting IRFs having the posterior means of λ_i of less than 0.6. Recall that $\lambda_i = 1$ for all IRFs in standard linear regression, and λ_i being much less than 1 indicates that the IRF's contribution to the analysis is downweighted. The smallest posterior mean of λ_i among the 714 IRFs in the analysis is 0.2, which implies that on average the variance of the IRF for which $\lambda_i = 0.2$ is inflated by a factor of $1/0.2 = 5$ relative to the variance for an IRF that is not unduly influential. In follow-up exploratory analyses, we did not find any clear reason why some IRFs were statistical outliers and therefore downweighted; even examining more up-to-date cost data than that available to Carter et al. (2002) yielded results that agreed with those presented here. The finding of non-normality of the regression errors in these data agrees with other authors who have found that cost data in health care studies are often quite non-normally distributed and that simply taking the logarithm of cost may not be sufficient to achieve normality (O'Hagan and Stevens 2003). The lack of a clear reason why certain points are statistical outliers also suggests the lack of a viable policy reason for excluding them from the analysis.

APPROPRIATENESS OF THE BAYESIAN OUTLIER ACCOMMODATION MODEL

Table 2 shows the R^2 value for the standard linear regression—column (a)—and for the Bayesian model—column (b). The posterior mean of R^2 in the Bayesian analysis is 0.532 as compared to its estimate of 0.465 for the standard linear regression, suggesting that the Bayesian model explains a greater proportion of the variation in the data than does standard linear regression. A more formal Bayesian test was done to confirm this observation. The first step of this test is to specify the prior odds of a model having a heavier tailed error distribution (i.e., a t -distribution with ν being less than 30 degrees of freedom) versus a model that assumes approximately normality of the errors (ν greater than or equal to 30) to be equal to 1; this holds given prior specification (1) for ν . The posterior odds are then computed as the posterior probability of ν being less than 30 divided by the posterior probability that ν is greater than or equal

to 30. The posterior odds in favor of v being less than 30 is estimated to be 3,124. Since the prior odds are equal to 1, the posterior odds equal the *Bayes factor*. The Bayes factor indicates the extent to which the data change the prior odds to yield the posterior odds. It is a measure that is frequently used for assessing the adequacy of one statistical model versus another in Bayesian analysis. Bayes factors of 150 and greater are typically considered to be very strong evidence in favor of the model (Kass and Raftery 1995), so a Bayes factor of 3,124 is overwhelming evidence in favor of the Bayesian outlier accommodation model over a model that assumes approximately normally distributed errors.

RESULTS: PAYMENT REGRESSION

Table 3(b) shows the results for the payment regression using the Bayesian outlier accommodation model. Both the rural and the low-income patient adjustments are slightly lower than in the standard linear regression model—Table 3(a). The posterior distribution for v is similar here as for the fully specified regression, with mean 8.4 and 95 percent posterior probability interval (4.8, 16.4). There are 24 statistical outlier IRFs identified in the payment model. Eighteen of these 24 IRFs are also identified as being statistical outliers in the fully specified regression presented in Table 2(b). The correlation of all of the posterior means of λ_i for the fully specified and payment regressions is 0.92. The consistency between the set of IRFs that are identified as statistical outliers in both the payment and fully specified regressions suggests that the nonpayment variables appearing only in the fully specified regression do not explain why these IRFs are statistical outliers.

Given that the indirect teaching measure showed up as significant in the Bayesian outlier accommodation model but not in the linear regression, we revisited the payment regression by including the indirect teaching variable. The indirect teaching coefficient is not significant in the Bayesian outlier accommodation payment model (posterior mean 0.219, 95 percent posterior probability interval of $[-0.173, 0.605]$) nor in the standard linear regression model (mean 0.131, 95 percent confidence interval $[-0.275, 0.538]$) (Table 4).

If the Bayesian outlier accommodation model had been used to create payment adjustments, the payment adjustment for rural would have been 17.9 percent ($\exp[0.165] = 1.179$) rather than 19.1 percent. A lower adjustment would have resulted for the low-income percentage variable as well. For an IRF with a low-income patient ratio (DSH) of 0.20, the payment increase

Table 4: Payment Regression Parameters Using the (a) Ordinary Least Squares Model and the (b) Bayesian Outlier Accommodation Model When the Indirect Teaching Variable Is Included

<i>Variable</i>	<i>(a) Ordinary Least Squares (Mean and 95% Confidence Interval)</i>			<i>(b) Bayesian Outlier Accommodation Model (Mean and 95% Probability Interval)</i>		
	<i>Mean</i>	<i>2.5%</i>	<i>97.5%</i>	<i>Mean</i>	<i>2.5%</i>	<i>97.5%</i>
Low income	0.464	0.227	0.701	0.390	0.157	0.624
Low-income imputation	0.023	− 0.016	0.062	0.029	− 0.010	0.068
Rural IRF	0.177	0.112	0.242	0.167	0.103	0.231
Indirect teaching	0.131	− 0.275	0.538	0.219	− 0.173	0.605
Intercept	9.318	9.286	9.350	9.325	9.292	9.358
R^2		0.061			0.066	

under the current system is 9.2 percent (since $[1+0.20]^{0.484} = 1.092$) and it would have been 8.0 percent (since $[1+0.20]^{0.424} = 1.080$) had the Bayesian outlier accommodation model been used. Similarly, for the mean DSH value of 0.112 from Table 1, the difference would have been a payment increase of 4.6 percent using the Bayesian outlier accommodation model versus the current adjustment of 5.3 percent.

STABILITY OF COEFFICIENT ESTIMATES ACROSS APPROACHES

We conducted a simulation study to compare the stability of coefficient estimates across standard linear regression, the Bayesian outlier accommodation model, and the robust regression procedure as implemented by “rreg” in *Stata*³ when statistical outliers are present in the data. The first step was to create a dataset that was free of statistical outliers so that statistical outliers could be randomly added to the data for the simulation study; we created this dataset by trimming the 23 most extreme observations in the dataset such that these observations were no longer outliers.⁴ We then randomly redistributed the trimmed portions of these 23 statistical outliers to other observations in the dataset and fit the fully specified regression model, repeating this procedure 50 times. Table 5 summarizes the results of this simulation study for each of the regression coefficients. The second and third columns show the percent increase in the regression coefficients across the 50 simulations relative to the Bayesian outlier accommodation model for the standard linear regression

Table 5: Simulation Study: Variation in Regression Coefficients across Fifty Simulated Datasets for the Standard Linear (OLS) and Robust Regressions (RREG) versus the Bayesian Outlier Accommodation Model (BAYES)

<i>Variable</i>	<i>Percent Increase in Variance Relative to BAYES</i>		<i>Range of Regression Coefficients</i>		
	<i>OLS</i>	<i>RREG</i>	<i>OLS</i>	<i>RREG</i>	<i>BAYES</i>
Log(CMI)	263	47	0.209	0.119	0.094
Wage index	77	19	0.273	0.208	0.186
Low income	230	93	0.290	0.240	0.211
Low-income imputation	288	73	0.059	0.036	0.029
Indirect teaching	124	51	0.616	0.637	0.486
Freestanding IRF	166	2	0.069	0.032	0.034
Small IRF	70	21	0.045	0.040	0.035
Large IRF	616	30	0.084	0.036	0.032
Old certification date	229	0	0.072	0.048	0.049
New certification date	312	46	0.045	0.030	0.025
Large urban location	624	25	0.112	0.033	0.033
Rural location	127	20	0.086	0.064	0.060
Proprietary IRF	241	35	0.061	0.046	0.035

model (OLS) and the robust regression model (RREG). Both OLS and RREG clearly yield less stable regression coefficients across the 50 simulations than the Bayesian outlier accommodation model; the variance of the regression coefficient for log(CMI) for OLS is 263 percent larger than that for the Bayesian outlier accommodation model, while the corresponding figure for RREG (47 percent) is lower but still indicative of relative instability. The RREG always does much better than OLS, though not as well as the Bayesian outlier accommodation model. The ranges of the coefficient estimates across the simulations (last three columns of Table 5) confirm this trend, where the OLS range of estimates is always larger than that obtained by RREG or the Bayesian outlier accommodation model in all but one case (the indirect teaching coefficient has a larger range for OLS than for RREG), while RREG yields a larger range of estimates than the Bayesian outlier accommodation model in all but three instances.

DISCUSSION

Standard linear regression has been widely used for setting payment rates for prospective payment systems. The importance of obtaining regression

parameter estimates that are robust to statistical outlier IRFs is clear given the size of PPS expenditures; a shift from 19.1 percent to 17.9 percent for the rural adjustment would result in a redistribution of \$3.5 million to nonrural IRFs. It is important that the best possible approach be used to derive payment adjustments and to confirm that existing standard approaches yield reasonable results. Our simulation study showed that the coefficient estimates obtained using the Bayesian outlier accommodation model are less sensitive to extreme observations than either the standard linear or robust regression approaches, thus making it less likely that a policy analyst will be misled by statistical outliers. The Bayesian outlier accommodation model provides at the very least a useful approach to assess the sensitivity of parameter estimates to the normality assumption and statistical outliers. Skewness and non-normality are commonly found in cost data, and linear regression coefficient estimates can be extremely sensitive to departures from normality (O'Hagan and Stevens 2003). While the standard practice of specifying the average cost per case outcome variable in the logged form is helpful for determining payment adjusters, our analysis shows that this may not fully solve the problem of developing the most suitable model in the context of developing facility payment adjustments in prospective payment systems. O'Hagan and Stevens (2003) argue that using methods that rely on asymptotic normality for the analysis of case-level health care cost data may be nonrobust if the data are highly non-normal; our analysis shows that non-normality of cost data can also be an issue at the facility level. The Bayesian outlier accommodation model fit here does not rely on assumptions about asymptotically normally distributed errors.

The Bayesian outlier accommodation model retains the basic structure and interpretability of the commonly used standard linear regression approach, which should be appealing to practitioners who have used standard linear regression for decades to derive facility payment adjustments. We have also shown that the Bayesian framework can be used to formally test hypotheses about the suitability of various models as part of the analysis; in this case, we showed that the data overwhelmingly support the use of the Bayesian outlier accommodation model.

NOTES

1. This contrasts with non-Bayesian analyses in which the parameters are regarded as fixed (i.e., not random).
2. Prior beliefs can include ignorance about the parameter by specifying a noninformative prior distribution for that parameter (i.e., a prior distribution that does not favor certain parameter values over others).

3. As discussed above, we could not use *Stata's* *reg* procedure for the fully specified regression because *reg* does not accommodate facility weights. However, if *reg* is applied to a dataset that contains n_i duplicates of the observation for each hospital i , then the regression coefficient estimates will be correct (though standard error estimates will be incorrect). The *reg* procedure applied in this way rounds the facility weights to be integers, but the impact of this difference on the simulation study results is negligible.
4. We substituted $\lambda_i n_i$ as the case weight for the 23 extreme observations identified by the Bayesian outlier accommodation model (i.e., posterior mean of $\lambda_i > 0.6$) and fit a standard linear regression model using the modified weights. We then assigned "non-outlier" values for the outcome variable for these 23 extreme observations to be the fitted values from this regression.

REFERENCES

- Carter, G. M., M. Beeuwkes Buntin, O. Hayden, J. H. Kawata, S. M. Paddock, D. A. Relles, G. K. Ridgeway, M. Totten, and B. O. Wynn. 2002. *Analyses for the Initial Implementation of the Inpatient Rehabilitation Facility Prospective Payment System*. Report No. MR-1500-CMS. Santa Monica, CA: RAND.
- Dalton, K., and E. C. Norton. 2000. "Revisiting Rogowski and Newhouse on the Indirect Costs of Teaching: A Note on Functional Form and Retransformation in Medicare's Payment Formulas." *Journal of Health Economics* 19 (6): 1027–46.
- DeGroot, M. H. 1970. *Optimal Statistical Decisions*. New York: McGraw-Hill.
- Fryback, D. G., N. K. Stout, and M. A. Rosenberg. 2001. "An Elementary Introduction to Bayesian Computing using WinBUGS." *International Journal of Technology Assessment in Health Care* 17 (1): 96–113.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 1995. *Bayesian Data Analysis*. London and New York: Chapman & Hall.
- Hamilton, L. C. 1998. *Statistics with Stata 5*. Pacific Grove, CA: Duxbury Press.
- Harrell, F. E., and Y-C. T. Shih. 2001. "Using Full Probability Models to Compute Probabilities of Actual Interest to Decision Makers." *International Journal of Technology Assessment in Health Care* 17 (1): 17–26.
- Kass, R. E., and A. E. Raftery. 1995. "Bayes Factors." *Journal of the American Statistical Association* 90 (430): 773–95.
- Lave, J. R. 1984. "Hospital Reimbursement under Medicare." *Milbank Memorial Fund Quarterly* 62 (2): 251–68.
- Newhouse, J. P., S. Cretin, and C. J. Witsberger. 1989. "Predicting Hospital Accounting Costs." *Health Care Financing Review* 11 (1): 25–33.
- O'Hagan, A., and J. W. Stevens. 2003. "Assessing and Comparing Costs: How Robust Are the Bootstrap and Methods Based on Asymptotic Normality?" *Health Economics* 12 (1): 33–49.
- Rogowski, J. A., and J. P. Newhouse. 1992. "Estimating the Indirect Costs of Teaching." *Journal of Health Economics* 11 (2): 153–71.

- Sheingold, S. H. 1990. "Alternatives for Using Multivariate Regression to Adjust Prospective Payment Rates." *Health Care Financing Review* 11 (3): 31–41.
- Spiegelhalter, D. J., J. P. Myles, D. R. Jones, and K. R. Abrams. 1999. "An Introduction to Bayesian Methods in Health Technology Assessment." *British Medical Journal* 319 (7208): 508–12.
- Spiegelhalter, D., A. Thomas, N. Best, and W. Gilks. 1995. *BUGS Examples* (version 0.5), vol. 1. Cambridge, UK: MRC Biostatistics Unit.
- Spiegelhalter, D. J., A. Thomas, and N. G. Best. 2000. *WinBUGS* (version 1.3) statistical software. Cambridge, UK: MRC Biostatistics Unit.
- Stata Corporation. 2003. *Intercooled Stata* (version 8.0). College Station, TX: Stata Corporation.
- Uniform Data System for Medical Rehabilitation (UDSmr). 1997. *The Guide for the Uniform Data Set for Medical Rehabilitation* (Including the FIM™ Instrument) (version 5.1). Buffalo, NY: UDSmr.
- Verdinelli, I., and L. Wasserman. 1991. "Bayesian Analysis of Outlier Problems Using the Gibbs Sampler." *Statistics and Computing* 1: 105–17.
- Weisberg, S. 1985. *Applied Linear Regression*. New York and Chichester, UK: Wiley.