Issues for the Next Generation of Health Care Cost Analyses

Anirban Basu, PhD,† and Willard G. Manning, PhD‡*

Background: Given the characteristics of health care expenditure/cost data–a mass of observations at zero, and skewed positive expenditures, various alternative estimators have been developed that can address the analytical issues these characteristics raise. The field continues to develop new approaches and to evaluate the performance of the existing ones.

Objectives: We discuss the strengths and limitations in existing methods for estimation and for model specification and checking. We suggest some areas that need fuller development or a better understanding of how the estimation approach performs when the outcome exhibits the skewness and heavy right tails that are typical of health care data. We also address various other aspects of cost analysis that include dealing with induced censoring, estimating casual effects, and generating reliable predictions that may apply to many studies.

Results: No current method is optimal or dominant for all cost applications. Many of the diagnostics used in choosing among alternatives have limitations that need more careful study. Several avenues in modeling cost data remain unexplored.

Conclusions: Taken together, we hope that this essay would serve as a guide to the choice among methods and to the next generation of methodological research in this field.

Key Words: health care costs, skewness, transformation, generalized linear models, censored costs

(*Med Care* 2009;47: S109–S114)

From the *Department of Medicine, Center for Health and the Social Sciences, University of Chicago, Chicago, Illinois; †The National Bureau of Economic Research, Cambridge, Massachusetts; and ‡Department of Health Studies, Harris School of Public Policy Studies, University of Chicago, Chicago, Illinois.

Supported by the National Cancer Institute and the Agency for Healthcare Research and Quality, as well as from the Harris School of Public Policy Studies and from the Department of Medicine, both at the University of Chicago.

The opinions expressed here are not those of the University of Chicago, the National Bureau of Economic Research, or the National Cancer Institute. Any remaining errors are those of the authors.

Reprints: Anirban Basu, PhD, Department of Medicine, Center for Health and the Social Sciences, University of Chicago and The National Bureau of Economic Research, Cambridge, MA 5841 S. Maryland Ave, MC-2007, Chicago, IL 60637. E-mail: abasu@medicine.bsd.uchicago.edu.

ISSN: 0025-7079/09/4700-0109

There has been a substantial literature written since the 1970s that addresses the statistical issues in dealing with health care costs and other outcomes, with much of the focus on the issue of skewness in the outcome data and its implications for the choice of an appropriate nonlinear method for covariate adjustment.[1–5] The fundamental problems for empirical analysis lie with the distributional nature of the cost data–limited dependent variables with a substantial fraction of zeroes; skewed to the right; responses that may vary over the range of costs or other inherent nonlinearities (eg, the role of comorbidities). Therefore, standard methods (such as, simple linear models with dollars as the dependent variable) that are widely used to answer questions involving other economic outcomes need to be modified to account for these common characteristics of health care cost or expenditure data.

Historically, this line of work has had some unintended consequences, because addressing one statistical issue has created additional statistical problems. In the discussion, we mention 2 of these cases. Some of the issues that we raise are included because of concerns that current proposals for estimators may have hidden consequences that need to be discovered. We will forgo a full review of the current state of the econometrics of cost analysis; the articles by Huang[6] and Mullahy[7] in this volume and their references provide a much more detailed view of the relevant literature for models that examine either survival-based or conditional mean models. We devote most of our attention to identifying issues in cost analyses that we think have not received adequate attention thus far and those that could lead to important research topics in the near future in the area of cost analyses. Although much of our discussion on issues associated with nonlinear adjustments models cuts across all types of analyses, we will also give separate attention to issues related to estimation of causal effects and predictions.

BACKGROUND

The predominant goal of cost modeling has been to draw inferences based on the mean costs [ie, $E(Y|x) = \mu(X)$] and how different covariates affect mean costs primarily due to its direct relation to either total or incremental budgetary impacts. Such impacts motivated the cost analysis in the first place in so many cases. Although, other moments of the cost distribution can also be of interests,[8] we will mainly focus of issues that arise in the context of mean-based analyses.

Traditional Single-Equation Models

Traditional linear regression usually fails to model consistently and reliably the mean of a skewed distribution with a heavy right tail because of the nonlinearity in the response, the instability caused by skewness and kurtosis, and/or the inefficiency due to the common failure to deal with the heteroscedasticity (variance increases with the mean). The field has considered 2 major classes of alternatives to least squares to address these issues with one based on transformation of the outcome and the other based on generalized linear models (GLM). Ordinary least squares (OLS) regression of logarithmic or MLE for Box-Cox transformations of Y on covariates X can overcome the skewness and may reduce problems of heteroscedasticity and kurtosis,[9] but does not result in a model for the mean $\mu(X)$ on the original dollar scale, a scale that in most applications is the scale of interest.[10,11] To draw consistent inferences about the mean $\mu(X)$ or any functional thereof in the natural scale of Y, complicated retransformations may be necessary, especially if the variances (or higher order moments) differ by clinical subgroup, treatment group, or by policy status.[1,11] To avoid the potential bias problems in retransformation, the use of GLM is increasingly becoming popular,[12] where a link function directly relates $\mu(x)$ to a linear specification $x^{\tau}\beta$ of covariates, including possible polynomials in covariates and interactions as needed. The retransformation problem is eliminated by transforming $\mu(x)$ instead of transforming the outcome variable Y. For example, if the response is proportional to the covariates, then the mean of Y given x is $\mu(x)^{x}\tau^{\beta}$. The link is the natural log, because $\log \mu(x) = x^{\tau}\beta$. Although the log link is ubiquitous, there is often no theoretical guidance as to what should be the appropriate link function or the variance function for the data at hand. The most common approach has been to rely on a series of diagnostic tests for candidate link and variance function models.[5,13]

But one of the overwhelming issues that often arises in this literature is that the analyst faces a dilemma in the choice between a potentially asymptotically biased (inconsistent) model based on transformation of Y or a consistent but potentially much less efficient method based on GLM approaches.[4] Even in the class of GLMs, the usual specification search may not necessarily generate efficient choices.

Advanced Single-Equation Models

Several new models have been proposed that overcome some of these specification problems for the GLM formulation. We proposed regression using a generalized gamma distribution that encompasses several other distributions as special cases.[14] Although the implicit link function is still a log link, the generalized gamma regression offers a flexible way to model a variety of shapes for the cost distribution, and therefore provides robustness to the mean estimates, as well as a test of whether the data are better approximated by some GLM versus log-based least squares formulations. Basu and Rathouz developed an extension of the traditional GLM where they use a Box-Cox style power link function that is directly estimated from the data so as to offer a more flexible functional form to the mean estimator.[15] Their model estimates the mean-variance parameters, jointly with those for the mean function, to capture the underlying heteroscedasticity and to overcome some of the efficiency limitation of traditional GLMs.[3,4] By allowing for link functions and variance functions other than the canonical ones, the Basu-Rathouz approach is less prone to misspecification and is potentially much more efficient than the usual approach. Other semi-parametric approaches have also been proposed in the literature based on modeling the hazard of cost accumulation.[16–18]

Going Beyond Single-Equation Models

All of the earlier estimators that we discussed thus far relate a central tendency of the response variable to covariates using a single, often highly nonlinear, functional form. The covariate effect is not formally (or directly) allowed to vary by the levels of the response, although different parts of the distribution of cost can have differential responses to a covariate that are brought about indirectly via other covariates in the model. Estimators which relax this assumption have also been proposed. They allow for different parts of the cost distribution to have different responses to covariates by using different parameters or functional forms for each part. That is, these estimators allow for the relationship between the response and the covariates to vary directly with the levels of the response.

The most common way to achieve this type of heterogeneity in effect is via 2-part models where the probability of incurring any costs is modeled separately from the level of costs given any health care utilization.[1,5,19] Duan et al[1] proposed an extension of 2-part models that allows inpatient and outpatient care to have different responses to covariates. A generalized version of multipart models was specified by Gilleskie and Mroz,[20] who have suggested that one can use a series of conditional models to address the skewed nature of health care expenditures. Building on the work by Efron[21] and Donald et al,[22] they suggest breaking up the dependent variable into different segments, modeling the probability p of being in a specific segment as a function of the covariates x's, and then using means conditional on being in each segment.

A potentially more parsimonious model that can be used to partition the distribution based on latent characteristics of patients is proposed by Deb and Trivedi.[23] Using latent class models, they model the distribution of health care cost as a mixture of 2 distributions–one for infrequent users and the other for frequent users. This is in contrast to the multipart models from the Health Insurance Experiment, where the classes are observed, not latent.[19]

In parallel to the burgeoning of new estimators for cost data, there were also a host of studies that compared many of these alternative models head-to-head using simulations and also real application data.[1,4,13,18,24,25]

Despite the variation in the sources of data analyzed in these works, one consensus always seems to reaffirm itself–there is not one universally optimal estimator for costs data. Instead, one must pay close attention to the nature and the distribution of the specific cost data that the analyst has at hand to select a handful of alternatives and then apply rigorous goodness-of-fit tests to select the most appropriate model for that data set. It is with this notion that we look

forward to highlight some areas in this field that could benefit substantially with further research.

ISSUES IN THE ANALYSES OF COST DATA THAT CUT ACROSS ALL PURPOSES OF ANALYSIS

The Longevity of OLS Regressions

Despite a large literature now suggesting the utility of nonlinear models for cost data, OLS regression is still a popular choice for many applied researchers. For some cost data, and depending on the specification for the covariates used, the identity link (untransformed y) could very well be the optimal link function.[26] However, the use of OLS seems to flourish mostly due to statistical philosophy (eg, works by the developers of the Diagnostic Cost Group [DxCG group], where patients were categorized into finely defined diagnostic bins to enhance predictive power[27–29]). In a number of these studies, the models are inherently nonlinear in the covariates because of the inclusion of interaction terms. In such contexts, there are a few questions that are still open.

1. When is OLS, with a variety of interactions and polynomials of covariates (to deal with nonlinearities in the response to covariates) adequate, in terms of robustness or reducing the risk of overfitting, to use with cost data?
2. What sample sizes are necessary to no longer be concerned about the robustness of key parameter estimates–to be able to employ the DxCG or similar approaches without worrying about the frailty of estimates for small- to moderate-sized subgroups/categories in the data? Does robustness depend on the rareness of important characteristics (eg, rare diseases or subgroups), degree of skewness, fraction of zeroes, etc.? Clearly, samples of the order of 30 million Medicare enrollees would be well behaved using least square if one was careful about checking the model's linearity or are not concerned about a few very thin cells, such as the one that occurred with Version 6 of the DxCG approach for a relatively rare but very expensive population to treat (ESRD/diabetes/complications). Is a sample of 500 to 1000 too small? Where in between does concerns about robustness (insensitivity to deviant cases) versus the risk of overfitting influential outliers (that may occur in a subsample) drive modeling decisions?
3. What standards should be used for assessing the goodness of fit for very large data sets, compared with those for more modest sample sizes?

Issues With GLMs and Their Extensions

Over the last 10 years, quite a lot of attention has been paid to the estimation of conditional mean models based on a class broadly referred to as GLM (McCullagh and Nelder, 1989). Many of the questions earlier also apply to the GLM approaches. In addition, other questions still remain about the performance of these models.

1. Given that some GLM models can deal internally with zeroes (as can happen with some Box-Cox transforms, such as the ½ power, ¼ power, and 1/[an even integer >0]), when do one need a 2-part GLM versus a 1-part GLM? Mullahy[3,7] considers such modeling unnecessary relative to single-equation methods, but there are applications where the fraction of zeroes are substantial enough to suggest either a 2-part model or substantial attention to a highly nonlinear specification of the covariates to deal with the likely failure of a single-equation model with a simpler index function, $x\beta$.
2. Because estimates based on some GLMs can be consistent but very imprecise for certain data generating mechanisms (especially for either high log-scale error variances, or heavy log-scale tails, which is typical of inpatient costs),[4] there needs to be more robust, consistent, and efficient methods developed for this subclass of data generating processes. Alternatives beyond our modification of the generalized gamma distribution[14] or dealing with alternative links than its implied choice of the log in the generalized gamma model can prove to be quite valuable.
3. There needs to be more work on less parametric alternatives and when to choose them, beyond the initial work by Gilleskie and Mroz.[20] Their proposal shows great promise but has seen very little systematic evaluation beyond their initial article. Their model is based on: (1) dividing the range of the dependent variable into segments for ranges of costs (0, 1–100, 101, . . .); (2) modeling how the choice of segments depends on covariates using a discrete hazard model; (3) estimating the mean expenditure per segment; and (4) estimating the overall response as the sum of the probabilities for each segment multiplied by its conditional mean. What are the trade offs in the number of segments (bins, in their original terminology) used? Within segment, would there be any gains to adjusting for covariates, x, especially in the upper bin where the largest fraction of spending occurs?

Issues With Diagnostic Tests for Determining Optimal Model

To date, most researchers have employed a set of diagnostics that were developed either for the least squares model or some of the GLM (such as the logit model). There needs to be a far more systematic critical review of diagnostics focusing on when the diagnostics work reliably, or efficiently, or not for data such as health care costs. Using diagnostics that work from raw-scale residuals is potentially quite problematic for data sets with small to moderate sample sizes (N′s <5 digits?) because these tests may not be very robust to the distributional characteristics (skewness, heavy right tail) that lead so many analysts to abandon simple least squares as our main analytical approach.

We note 3 examples from our own experience that can be generically applied to all the estimators we have discussed as they are based on raw-scale residuals. First, the modified Hosmer-Lemeshow test,[30] which identifies systematic patterns in mean raw-scale residuals across deciles of the linear predictor, is particularly prone to influential outliers and is imprecise because of the number of degrees of freedom involved in the test, typically 5 or 10. Second, the Link and RESET tests,[31] which aims to identify misspecification in functional forms, are more parsimonious tests but can provide both false positives and false negatives as a result of mildly influential observations for the original analysis; the RESET is particularly susceptible. Third, some of the split-sample cross-validation approaches, which try to identify degrees of overfitting, can be poorly behaved when

dealing with data as skewed and heavy-tailed as health care cost. There is a robustness issue when using the cross-validation Copas[32] tests when there is extreme skewness and potentially influential outliers. One version of the test involves estimating the model on a random half of the data, forecasting the results to a second half of the data or test sample, and testing if response of y to the forecast has a slope significantly different from one. Any significant or appreciable departure from one indicates that the estimator overfitted the original data and cannot forecast well to random samples from the same population. The high-end or catastrophic cases are potentially problematic for this approach. In data such as health expenditures, a highly influential "catastrophic" case(s) is always in one sample or the other, estimation versus test subsamples. Either the initial estimate is subject to overfitting, or the forecast to the test sample will be subject to overfitting, or both.

One could consider a robust alternative to such a model, such as an alternative to the Davidson and McKinnon test on some appropriate scale to compare between nonnested models.[33] But doing so on the raw-scale raises the same concerns about the robustness of the test results in the presence of zeroes, skewness, and heavy right tails of the distribution. Specifically, should we use raw-scale versions or scale-of-estimation versions of the Copas Tests?

Issues With Censoring in Longitudinal Costs Data

Many observational databases and some clinical trials data suffer from censoring issues. In some of the earlier work, the analysis was limited to fixed width periods of observation. In others, the analysis was reduced to a per-member per-month calculation with weights equal to the number of months under observation. Some of the more recent work has highlighted the dependent nature of censoring in costs data[34–38]; See the articles by Huang[6] in this volume on survival-based methods. It is our sense that there are several questions that remain unanswered in dealing with censored costs.

1. Under the assumptions of random censoring, regression estimators that are weighted with inverse of the probability for surviving censoring can provide consistent estimates of treatment costs. Do such methods remain consistent if treatment has a direct effect on survival? If so, then what conditions have to apply?
2. How do we model cost-trajectories for different death cohorts? One of the empirical regularities across countries and across disease groups is that costs accelerate in the last 6 months or so of life.[39,40] See Brown et al for an illustration of this in breast cancer patients.[41] Lipscomb et al address this heterogeneity to some extent using a stratified Cox model with each month being a different stratum, to infer the "time shape" of costs post the incident disease event.[18] These approaches need to be more widely tested. We would encourage explorations of alternatives that are specifically designed to deal with the possibility that a case may be censored but still on a high cost death trajectory. The conjunction of acceleration in expenses toward the end of life and censoring is a difficult analytical problem.

Specific Issues in Cost Analyses Used for Evaluation

Under strong ignorability conditions (ie, the unobservable differences across observations are not correlated with the "treatment" or policy variable being studied and there are no issues of adverse selection), such as that which occurs in randomized trials and in natural experiments, there are a number of issues that need to be more fully examined.

1. Sample sizes in clinical trials are often powered to look at efficacy on clinical outcomes and not costs. Yet, there is a growing popularity of doing cost analysis alongside clinical trials. One example of this is the recently completed CATIE study that was powered to compare time to discontinuation between antipsychotic drugs.[42] Unfortunately, ex-post calculation reveal that the sample sizes employed had only a 10% power to look at cost-effectiveness outcome.

   Generally, it is hard to find any discriminatory power across alternative estimators of health care costs when sample sizes are small. One alternative may be to consider doing utilization-based analysis instead of a full cost analysis. Are we better off in terms of precision to model utilizations in these setting rather than costs or by a weighted utilization measure? The latter could be generated by multiplying utilizations with a constant unit cost estimates for all utilization (price per average visit) instead by prices for each type of utilization? This would reduce the overall variance by eliminating the patient-to-patient and visit-to-visit variation in the cost per encounter. Bayesian modeling may be an interesting approach to explore in this area, especially whether model averaging across alternative estimators can be done. As long as the treatment does not change the cost or content of an encounter or the mix of different types of encounters, there is no risk of bias in this approach. If either occurs, then a fuller analysis may be required, including microcosting.
2. How does this common unit cost approach affect variability in total costs or the estimates of the incremental/marginal cost of an intervention? How should differences across research sites, cities, or countries be addressed? If such an approach is to be employed, how should we establish standard unit cost estimates that can be used to weight utilizations?
3. Propensity scores (PS) are often used as an alternative to regression techniques for risk adjustments. In such methods, one estimates how various characteristics affect the probability of treatment receipt, creates a score based on this estimation and then compares observed outcomes between treated and untreated subjects conditional on this score (Rosenbaum and Rubin, 1983). The theory of PS suggests that, conditional on this scalar PS, all of the selection bias generated by differences in observed covariate values between the treatment and control group can be removed.[43] However, much of the existing work documenting the robustness of these methods is based on linear models. Whether these features of PS carry over to outcomes generated via nonlinear model are not obvious and require further investigations.[44]

When strong ignorability conditions are absent, as is the case in most observational data, there are a number of issues that require more examination.

4. All types of 2-stage methods for instrumental variables do not readily extend to nonlinear models. The 2-stage residual inclusion (2SRI) method seems to be a consistent alternative under some general assumptions. Consistent 2SRI methods for specific nonlinear models have been developed.[45–50] Wooldridge suggests the use of the 2SRI method for count data models.[51,52] Recently, Terza et al have shown the consistency of this methods for a broad class of nonlinear models.[53] In this method, the residuals from the first stage, where the endogenous variable is regressed on a vector of exogenous variables (X's) and instrumental variables (Z), are computed and then included as an additional covariate in the second-stage nonlinear regression of outcomes on the endogenous variable. More research on the robustness issues with 2SRI are needed with respect to the underlying assumptions. Further, there are a series of questions about which measure of residuals or deviances to use, what are the implications for model selection and specification searches? Furthermore, the comparisons between residual inclusion methods and control functions approaches proposed by Heckman and others need more investigation.
5. Distinction between local average treatment effects[54] (eg, the average treatment effect for a portion of the population that is induced to change treatment due to varying levels of an instrumental variable) and policy relevant treatment effect parameters such as the Average Treatment Effect, the Treatment effect on the Treated, and the Treatment effect for the Untreated are quite important for answering policy questions[55] but are often not made in the cost literature. (The exception to this seems to be the literature on the incremental effects of insuring the uninsured). The local instrumental variable (LIV) method is an attractive way to characterize these distinctions among effects.[56] More work is needed to develop LIV methods in nonlinear settings with the characteristics common to cost data.[55,56]

FUTURE DIRECTIONS IN METHODS USED FOR PREDICTION

We find this aspect of the work in the field still in its infancy. Several issues need attention.

1. To some researchers perhaps the most surprising aspect of this work is that none of the models developed thus far have been able to produce an R^2 or pseodu-R^2 of more than 30% in either in-sample or out-of-sample situations. The earlier work by Newhouse et al,[57] on maximum explainable variance suggests that time invariant factors should account for about 10% of the variance in inpatient expenditures and 30% in outpatient. How much of the discrepancy between these estimates and those in the literature are analytical methods? How much of this is weak measures of either case mix or severity conditional on case mix, weak in the sense of misclassification or measurement error?

   Investigators using concurrent measures of case mix often achieve better explanatory power. This gain is largely the result of regressing a variable on its components, a dubious practice. Another equally dubious practice is to use current measures of health status, case mix, or severity to explain, or postdict past costs.[58] How much did you spend on health care in 2008 and what is your health today?
2. What are the criteria for assessing optimal predictions? Do we need unbiased covariate estimates for prediction or are we satisfied with lower mean square error in out-of-sample data? How do we differentiate between noise and heterogeneity (the work by DxCG group addresses this to some extent). This raises important questions regarding the right loss function with which we can trade-off sensitivity and specificity of cost predictions.
3. When are multipart and higher dimensional nonlinear models more suitable for prediction, despite larger number of parameters estimated?
4. Risk prediction models in costs have been primarily developed for calibration (getting the mean correct). Less attention has been paid to develop models that could reliably predict the costliness of a particular individual. (We thank an anonymous reviewer for pointing this out).

GENERAL COMMENTS

The field has often gone from one method to another, only to find that the new method has serious limitations. To deal with skewness, transformed models were widely adopted. But estimation scale inferences were not sufficient, which raised the retransformation issue[10] and concern about bias from ignoring nonconstant variability across covariates.[3,11] Modeling mean functions directly (as in the GLM approach) was an alternative approach that avoided these 2 issues,[5] but led to a loss of precision in some extremely skewed data. In others, GLM performed better than transformation methods. Solving the retransformation bias problem led to methods, which are consistent but sometimes inefficient, losing some of the robustness gains of the first generation of models.[4]

There are a number of newer alternatives that Profs. Huang, Mullahy, and we have discussed. Their strengths and weaknesses need greater scrutiny. It is only by conducting such a statistical SWOT analysis that we can make more informed decisions about which econometric approach might best suit the analysis at hand.

Although the progress in addressing the statistical issues in cost analysis has been substantial over the last 40 years, there are still important issues that need attention. We hope that this set of directions for the near term will help to set the agenda for some of this work.

ACKNOWLEDGMENTS

The authors thank the editors and 2 anonymous reviewers for their comments.

REFERENCES

1. Duan N, Manning WG, Newhouse JP, et al. A comparison of alternative models for the demand for medical care. *J Bus Econ Stat*. 1983;1:115–126.
2. Jones AM. Health Econometrics. In: Culyer A, Newhouse J, eds. *Handbook of Health Economics*. Amsterdam, The Netherlands: Elsevier;

2000.
3. Mullahy J. Much ado about two: reconsidering retransformation and the two-part model in health econometrics. *J Health Econ*. 1998;17:247–281.
4. Manning WG, Mullahy J. Estimating log models: to transform or not to transform? *J Health Econ*. 2001;20:461–494.
5. Blough DK, Madden CW, Hornbrook MC. Modeling risk using generalized linear models. *J Health Econ*. 1999;18:153–171.
6. Huang Y. Cost analysis with censored data. 2009;47(suppl 7):S115–S119.
7. Mullahy J. Econometric modeling of healthcare costs and expenditures: A survey of analytical issues and related policy considerations. 2009; 47(suppl 7):S104–S108.
8. Vanness DJ, Mullahy J. Perspectives on mean-based evaluation of health care. In: Jones A, eds. *The Elgar Companion to Health Economics*. Cheltenham, United Kingdom: Edward Elgar Publishing; 2006.
9. Box GEP, Cox DR. An analysis of transformations. *J R Stat Soc Series B*. 1964;26:211–252.
10. Duan N. Smearing estimate: a nonparametric retransformation method. *J Am Stat Assoc*. 1983;78:605–610.
11. Manning WG. The logged dependent variable, heteroscedasticity, and the retransformation problem. *J Health Econ*. 1998;17:283–295.
12. Wedderburn RWM. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*. 1974;61:439–447.
13. Buntin MB, Zaslavsky AM. Too much ado about two-part models and transformation? Comparing methods of modeling medical expenditures. *J Health Econ*. 2004;23:525–542.
14. Manning WG, Basu A, Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes. " 2003 NBER Working Paper No. t0293." *J Health Econ*. 2005;24:465–488.
15. Basu A, Rathouz PJ. Estimating incremental and marginal effects on health outcomes using flexible link and variance function models. *Biostatistics*. 2005;6:93–109.
16. Basu A, Manning WG, Mullahy J. Comparing alternative models: log vs Cox proportional hazard? *Health Econ*. 2004;13:749–765.
17. Jain AK, Strawderman RL. Flexible hazard regression modeling for medical cost data. *Biostatistics*. 2000;3:101–118.
18. Lipscomb J, Ancukiewicz M, Parmigiani G, et al. Predicting the cost of illness: a comparison of alternative models applied to stroke. *Med Dec Making*. 1998;18(suppl 2):S39–S56.
19. Cragg J. Some statistical models for limited dependent variable with application to the demand for durable goods. *Econometrica*. 1971;39:829–844.
20. Gilleskie DB, Mroz TA. A flexible approach for estimating the effects of covariates on health expenditures. *J Health Econ*. 2004;23:39–418.
21. Efron B. Logistic regression, survival analysis, and the Kaplan-Meier curve. *J Am Stat Assoc*. 1988;83:414–425.
22. Donald SG, Green DA, Paarsch HJ. Differences in wage distributions between Canada and the United States: an application of a flexible estimator of distribution functions in the presence of covariates. *Rev Econ Stud*. 2000;67:609–633.
23. Deb P, Trivedi PK. The structure of demand for health care: latent versus two-part models. *J Health Econ*. 2002;21:601–625.
24. Baser O. Modeling transformed health care cost with unknown heteroskedasticity. *App Econ Res Bull*. 2007;1:1–6.
25. Basu A, Arondekar BV, Rathouz PJ. Scale of interest versus scale of estimation: comparing alternative estimators for the incremental costs of a comorbidity. *Health Econ*. 2006;15:1091–1107.
26. Diehr P, Yanez D, Ash A, et al. Methods for analyzing health care utilization and costs. *Ann Rev Pub Health*. 1999;20:125–144.
27. Ash AS, Ellis RP, Pope GC, et al. Using diagnosis to describe populations and predict costs. *Health Care Financ Rev*. 2000;21:7–28.
28. Pope GC, Ellis RP, Ash AS, et al. Principal inpatient diagnostic cost group model for Medicare risk adjustment. *Health Care Financ Rev*. 2000;21:93–118.
29. Zhao Y, Ash AS, Ellis RP, et al. Predicting pharmacy costs and other medical costs using diagnosis and drug claims. *Med Care*. 2005;43:34–43.
30. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. 2nd ed. New York, NY: John Wiley & Sons; 1995.
31. Pregibon D. Goodness of link tests for generalized linear models. *App Stat*. 1980;29:15–24.
32. Copas JB. Regression, prediction, and shrinkage. *J R Stat Soc Series B*. 1983;45:311–354.
33. Davidson R, McKinnon JG. Some non-nested hypothesis tests and the relations among them. *Econometrica*. 1982;49:551–565.
34. Lin DY, Feuer EJ, Etzioni R, et al. Estimating medical costs from incomplete follow-up data. *Biometrics*. 1997;53:113–128.
35. Lin DY. Linear regression analysis of censored medical costs. *Biostatistics*. 2000;1:35–47.
36. Bang H, Tsiatis AS. Estimating medical costs with censored data. *Biometrika*. 2000;87:329–343.
37. Lin DY. Proportional means regression for censored medical costs. *Biometrics*. 2000;56:775–778.
38. Lin DY. Regression analysis of incomplete medical cost data. *Stat Med*. 2003;22:1181–1200.
39. Scitovsky AA. The high cost of dying: what do the data show? *Milbank Mem Fund Q Health Soc*. 1984;62:591–608.
40. Stearns SC, Norton EC. Time to include time to death? The future of health care expenditure predictions. *Health Econ*. 2004;13:315–327.
41. Brown M, Riley G, Schussler N, et al. Estimating health care costs related to cancer treatment from SEER-Medicare data. *Med Care.* 2002;40(suppl 8):IV–104–IV–117.
42. Rosenheck RA, Leslie DL, Sindelar J, et al. Cost-effectiveness of second generation antipsychotics and perphenazine in a randomized trial of treatment for chronic schizophrenia. *Am J Psychiatry*. 2001;63:2080–2089.
43. Rosenbaum PR, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41–55.
44. Basu A, Polsky D, Manning WG. *Use of Propensity Scores in Non-Linear Response Models: The Case for Health Care Expenditures*. Cambridge, MA: National Bureau of Economic Research Working Paper Series, w14086; 2008.
45. Blundell RW, Smith RJ. Estimation in a class of simultaneous equation limited dependent variable models. *Rev Econ Stat*. 1989;56:37–58.
46. Blundell RW, Smith RJ. Simultaneous microeconmetric models with censored or qualitative dependent variables. In: Maddala GS, Rao CR, Vinod HD, eds. *Handbook of Statistics*. Vol 2. Amsterdam, The Netherlands: North Holland Publishers; 1993;117–1143.
47. Rivers D, Vuong QH. Limited information estimators and exogeneity tests for simultaneous probit models. *J Econom*. 1988;39:347–366.
48. Smith RJ, Blundell RW. An exogeneity test for a simultaneous equation tobit model with an application to labor supply. *Econometrica*. 1986; 54:679–685.
49. NeweyWK. Efficient estimation of limited dependent variable models with endogenous explanatory variables. *J Econom.* 1986;36:231–250.
50. Bhattacharya J, Goldman D, McCaffrey D. Estimating probit models with self-selected treatments. *Stat Med*. 2006;25:389–413.
51. Wooldridge JM. Quasi-likelihood methods for count data. In: Pesaran M, Schmidt P, eds. *Handbook of Applied Econometrics*. Vol II: Microeconometrics. Malden, MA: Blackwell Publishers Ltd; 1997.
52. Wooldridge JM. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, MA; 2002.
53. Terza JV, Basu A, Rathouz PJ. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *J Health Econ*. 2008;27:531–543.
54. Angrist J, Imbens G, Rubin D. Identification of causal effects using instrumental variables. *J Am Stat Assoc*. 1996;91:444–455.
55. Basu A, Heckman JJ, Navarro S, et al. The use of instrumental variables in the presence of heterogeneity and self-selection: an application in breast cancer patients. *Health Econ*. 2007;16:1133–1157.
56. Heckman JJ, Vytlacil E. Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proc Natl Acad Sci USA*. 1999;96:4730–4734.
57. Newhouse JP, Manning WG, Keeler EG, et al. Objective measures of health and prior utilization as adjusters for capitation rates. *Health Care Financ Rev*. 1989;10:41–54.
58. Manning WG, Newhouse JP, Ware JE. The status of health in demand estimation: or, beyond excellent, good, fair, and poor. In: Victor Fuchs, ed. *Economic Aspects of Health, National Bureau of Economic Research*. Chicago, IL: University of Chicago Press; 1982.

