

A Comparison of Methods for Analyzing Health-Related Quality-of-Life Measures

Peter C. Austin, PhD

Institute for Clinical Evaluative Sciences, North York, Ontario; Department of Public Health Sciences, University of Toronto, Toronto, Canada

ABSTRACT

Objectives: Self-reported health status is often measured using psychometric or utility indices that provide a score intended to summarize an individual's health. Measurements of health status can be subject to a ceiling effect. Frequently, researchers want to examine relationships between determinants of health and measures of health status. Regression methods that ignore the censoring in the health status measurement can produce biased coefficient estimates. The authors examine the performance of three different models for assessing the relationship between demographic characteristics and health status.

Methods: Three methods that allow one to analyze data subject to a ceiling effect are compared. The first model is the classic Tobit model. The second and third models are robust variants of the Tobit model: symmetrically trimmed least squares and censored least absolute deviations (Censored LAD) regression. These models were fit to data from the Canadian National Population Health

Survey. The results are compared to three models that ignore the presence of a ceiling effect.

Results: The Censored LAD model produced coefficient estimates that tended to be shrunk toward 0, compared to the other two models. The three models produced conflicting evidence on the effect of gender on health status. Similarly, the rate of decay in health status with increasing age differed across the three models. The Censored LAD model produced results very similar to median regression. Furthermore, the censored LAD model had the lowest prediction error in an independent validation dataset.

Conclusions: Our results highlight the need for careful consideration about how best to model variation in health status. Based upon our study, we recommend the use of Censored LAD regression.

Keywords: ceiling effect, health status, Health Utilities Index, statistical methods, Tobit model.

Introduction

Self-reported health status is often measured using psychometric or utility indices that provide a score intended to reflect a person's health. Scores on psychometric scales are usually converted to a percentage of the maximum possible score [1]. Health utility indices typically measure health on a scale from 0 (or a small negative number) to 1, where 0 indicates dead and 1 indicates perfect health [2].

An essential issue in interpreting a health status index is the meaning of the extreme values of the index. Ceiling values are intended to represent states of perfection. Defining perfection is inher-

ently problematic, leading to the likelihood that the defined boundary is silently exceeded. Although utilities are traditionally measured on a scale from 0 to 1, it is useful in this context to consider health states better than 1. For example, the ambulation scale of the Health Utilities Index Mark 3 (HUI3) assigns a maximum score to an individual who is "able to move around the neighborhood without difficulty and without walking equipment." [3] Thus, a person who can walk but is unable to run or engage in sports and other vigorous activities might (depending on the responses to the other questions) be classified as perfectly healthy. Feeny et al. [3], comment that the Health Utilities Index may be subject to a ceiling effect and note that one can add supranormal levels for descriptive purposes. An upper limit on health status scores allows no explicit room for a treatment effect if the individual is in perfect health at baseline. A further argument for accounting for the censoring present in

Address correspondence to: Dr. Peter Austin, Institute for Clinical Evaluative Sciences, G-160, 2075 Bayview Avenue, North York, Ontario M4N 3M5, Canada.
E-mail: peter.austin@ices.on.ca

health status scores is that it allows one to better estimate a treatment effect under these conditions.

In studying the relationships between health status and predictors of health status such as age, gender, or socioeconomic status, investigators frequently construct regression models to quantify how health status varies with changes in subject characteristics. When a ceiling effect is present, standard regression models ignore the censoring that has occurred among those individuals with a health status that lies above the threshold for perfect health. One approach is to treat the index as if it was censored, with a score of 1 (or 100% being the largest observable value). In this scenario, for a subject with an observed score of 1, all that is known is that the subject's true health status is at least 1. In a recent article, Austin et al. [4] proposed the use of the Tobit model for analyzing measures of health status. In a related commentary, Grootendorst [5] proposed two robust alternatives to maximum likelihood estimation of the Tobit model, both developed by Powell [6,7], the censored least absolute deviations (censored LAD) estimator, and the symmetrically trimmed least squares estimator.

Previous research has shown that the classic Tobit model performs poorly when the distributional assumptions of the model are not satisfied. Maddala [8] commented that in the presence of heteroscedasticity, the usual Tobit estimates are inconsistent, and that there is only limited information about the direction of the bias. Furthermore, Greene states that the Tobit estimators are inconsistent in the face of non-normality of the error term [9]. In studying Bayesian extensions of the Tobit model, Austin demonstrated that there was strong evidence in favor of heteroscedasticity, compared to homoscedasticity in a Tobit model relating the Health Utilities Index to subject characteristics [10].

The purpose of this paper is to compare the performance of the classic Tobit model with that of censored LAD and symmetrically trimmed least squares for analyzing the relationship between demographic characteristics and the Health Utilities Index Mark 3. Data from the National Population Health Survey (NPHS) [11] were used for the analysis.

Methods

Data Sources

The HUI Mark 3 (HUI3), developed by Torrance and colleagues [3,12], describes an individual's health in terms of eight attributes: vision, hearing, speech, mobility, dexterity, cognition, emotion, and

pain/discomfort. There are five or six levels per attribute, resulting in 972,000 unique health states. A multiplicative scoring function is used to calculate the index. Because some health states are considered worse than dead, the index can take values from -0.03 to 1. The HUI3 health states from the NPHS were valued using the provisional scoring system, based on the HUI2 utility weights.

Cross-sectional data from the 1994/95 NPHS [11] were used for the current study. The NPHS questionnaire included components on health status, use of health services, demographic and socioeconomic status. The target population for the NPHS was household residents in all provinces. Patients in hospitals, and residents of long-term care institutions were excluded from the NPHS. The total number of respondents was 17,626. We restricted our analysis to those individuals who were between 20 and 80 years of age. We assumed that those subjects with HUI scores equal to 1 are censored observations, and that there is actually a range of supranormal health states among those with a reported score of 1.

Age is reported in 5-year increments in the NPHS. For age to be treated as a continuous variable, we chose to represent age using the mid-point of the interval. For a measure of socioeconomic status, the derived income adequacy variable was used, which categorizes subjects into five levels according to income adequacy. The income adequacy variable derived by Statistics Canada as a function of household income and size of the household takes on the following levels: lowest income; lower middle income; middle income; upper middle income, and highest income. We used the highest level of income adequacy as the reference level, and used indicator variables to represent each of the four lower levels of income adequacy. The number of chronic medical conditions was obtained by summing the number of affirmative responses regarding specifically diagnosed chronic medical conditions.

Statistical Models

Classic Tobit model. The Tobit model [9,13,14] is a well-known econometric regression model used in the presence of censored data. Assume that the true model is given by the following equation:

$$(1) \quad Y_i^* = \alpha + \beta_1 X_{i1}^* + \dots + \beta_k X_{ki}^* + \epsilon_i^*,$$

where Y_i^* denotes the individual's true health status score. However, an individual with an observed

health status score of 1, has a true $Y_i^* \geq 1.0$. Therefore, the observed dependent variable is given by:

$$Y_i = Y_i^*, \text{ for } Y_i^* < 1, \text{ and}$$

$$(2) \quad Y_i = 1, \text{ for } Y_i^* \geq 1.0.$$

The actual estimated regression equation will then be

$$(3) \quad Y_i = \alpha + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \varepsilon_i.$$

The classic Tobit model assumes that the error terms are normally distributed with uniform variance, $\varepsilon_i \sim N(0, \sigma^2)$. Previous research has shown that the Tobit model performs poorly in the presence of either heteroscedasticity or non-normality [8,9].

Symmetrically trimmed least squares. Symmetrically trimmed least squares regression, an alternative to maximum likelihood estimation of the Tobit model, has been proposed by Powell [6] to address the poor performance of the Tobit model in the presence of heteroscedasticity. This method, relaxes the assumption of homoscedasticity, assuming only that the distribution of the error term is symmetric around 0. The symmetrically trimmed least squares model describes the change in mean health status associated with changes in subject characteristics.

Censored least absolute deviations (censored LAD). Censored LAD, proposed by Powell [7], is a robust alternative to maximum likelihood estimation for the Tobit model. Censored LAD regression requires weaker assumptions than the symmetrically trimmed least squares model. This model is consistent for a wide class of error distributions, and is robust for heteroscedasticity. Paarsch [15] demonstrated that this method performs well for moderate sample sizes when the error terms follow a Cauchy distribution. This method is a censored version of least absolute deviations (or median) regression, which estimates regression coefficients to minimize the sum of the absolute value of deviations from the regression line. This differs from ordinary least squares regression, which minimizes the sum of the squared deviations from the regression line. The classic Tobit model describes the association between mean health status and subject characteristics, whereas the censored LAD model describes the association between the median health status and subject characteristics.

Traditional Statistical Models

For comparative purposes, three other models were fit to the data. First, a linear regression model, estimated using ordinary least squares (OLS), was fit to the data. This model ignored the potential censoring that occurs at the ceiling, and treated the observed health status as the true health status. Second, a linear regression model, estimated using OLS was fit to those subjects for whom the observed HUI lay below the ceiling value. Third, a median regression model [9] was fit to the data, ignoring the potential censoring that occurred at the ceiling. Median regression describes the change in median health status with changes in subject characteristics, in contrast to OLS regression, which describes the changes in mean health status with changes in subject characteristics.

Model Estimation

In this study, we used the iterative algorithms described by Johnston and DiNardo [14] to compute the symmetrically trimmed least squares and censored LAD models. For each of the two robust methods and for the median regression model, 95% confidence intervals and significance levels were computed using bootstrap methods [16]. For the classic Tobit model and the two OLS models, p -values and confidence intervals were obtained using model-based standard errors. The classic Tobit model and the two OLS models were estimated using SAS (2001, v 8.2) [17], whereas the other three models were estimated using S-Plus (1999, v 5.1) [18].

To determine predictive accuracy of each model, the data were randomly divided into equal sized derivation and validation sets. Each model was estimated using the derivation dataset. Using the coefficients derived from the derivation dataset, predicted HUI scores were computed for each subject in the validation dataset. For those subjects in the validation dataset with an observed HUI score of less than 1, the absolute prediction error was computed as the absolute value of the difference between the observed and predicted HUI scores. The median, first and third quartiles of absolute prediction error were calculated for each model.

In each model, it was assumed that the subject's health status, as measured using the HUI3, was related to age, gender, socioeconomic status, and number of reported chronic medical conditions. Both age and the square of age were entered into each model. By allowing a quadratic effect due to age, we allow for the possibility that health status

decays more rapidly with increasing age. Socioeconomic status was treated as a categorical variable, with the highest level of income adequacy being the reference level.

Results

The study cohort consisted of 14,460 subjects. Twenty-two percent of the subjects had an observed HUI score of 1. The median HUI score was 0.947. The first and third quartiles were 0.877 and 0.947, respectively. Figure 1 illustrates the distribution of HUI scores in the study population, and indicates that the distribution was decidedly non-normal and skewed. The R^2 values for models estimated using OLS on the entire data and on the nonlimit observations were 19.7% and 16.9%, respectively. Table 1 summarizes the coefficient estimates and 95% confidence intervals for the three models that incorporated the censoring present in the HUI score. Table 2 summarizes the coefficient estimates and 95% confidence intervals for the three models that ignored the censoring present in the HUI score.

In each of the six models, health-related quality of life (HRQL) decreased with decreasing income adequacy. Using the censored LAD model and the median regression model, one would infer that HRQL was not significantly different in the fourth income adequacy category than in the highest income adequacy category. Using the other four models, one would infer that HRQL was significantly lower in each income category than in the highest income category. The effect of income adequacy was more pronounced for the classic Tobit model than for the other two Tobit models. Similarly, the effect of income adequacy was more pronounced for the symmetrically trimmed least squares model than for the censored LAD model. In all six models, HRQL decreases as the number of chronic conditions increases. The magnitude of the effect of the number of chronic conditions is approximately 15% smaller in the symmetrically trimmed least squares model than in the classic Tobit model. The magnitude of the effect of the number of chronic conditions is approximately 30% smaller in the censored LAD model than in the

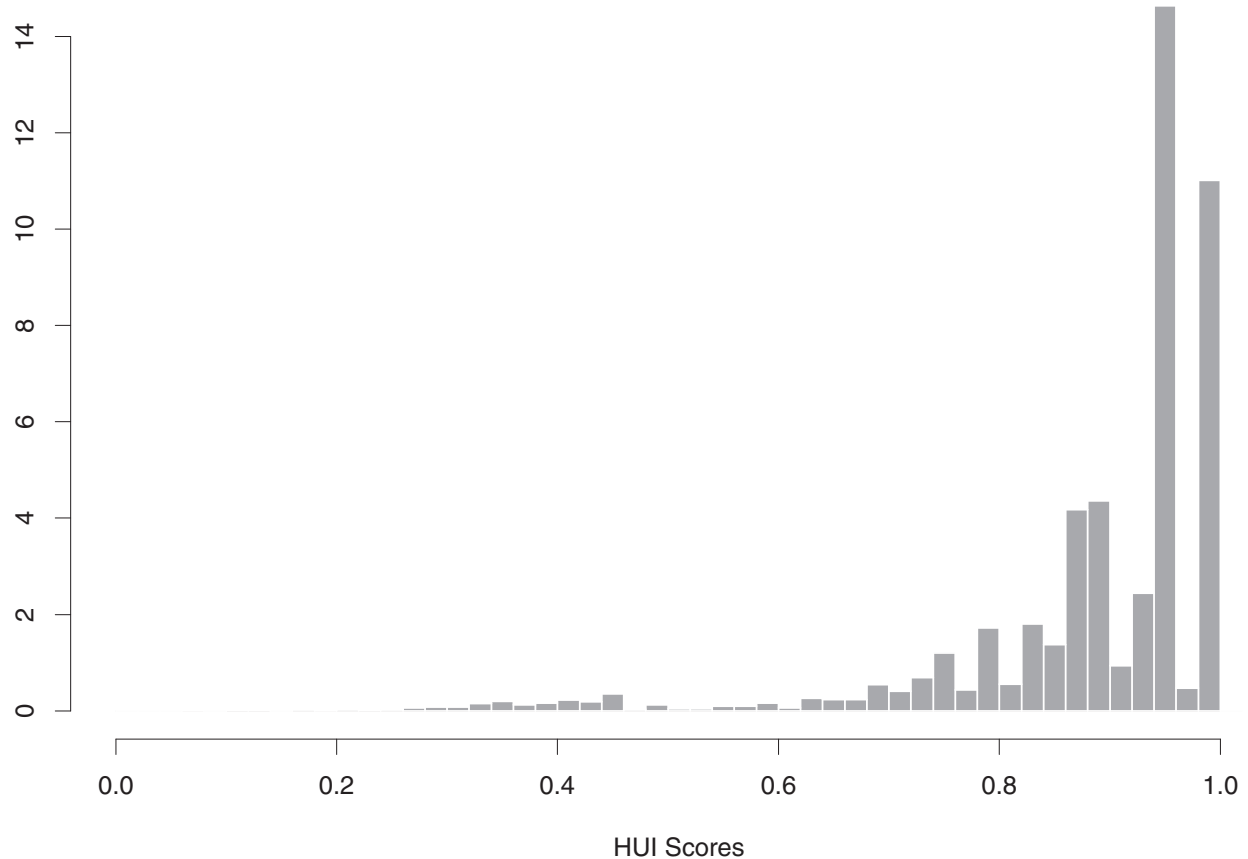


Figure 1 Distribution of HUI scores.

Table 1 Coefficient estimates for each of the three models that incorporate censoring in the HUI score

Coefficient	Classical Tobit model	Symmetrically trimmed least squares	Censored least absolute deviations
Intercept	1.095100 (1.071314, 1.118881)	1.085229 (< 0.001) (1.04920, 1.13485)	0.988816 (< 0.001) (0.95476, 1.00679)
Age	-0.003575 (< 0.0001) (-0.004565, -0.002585)	-0.003913 (< 0.001) (-0.005757, -0.002582)	-0.001217 (0.068) (-0.001906, 0.000132)
Age ²	0.000020 (0.0001) (0.000010, 0.000030)	0.000027 (< 0.001) (0.000014, 0.000045)	0.000007 (0.326) (-0.000007, 0.000015)
Male	-0.000222 (0.9303) (-0.005193, 0.004749)	0.001116 (0.644) (-0.003191, 0.005759)	0.002636 (0.286) (-0.000151, 0.003622)
Income ^{1*}	-0.070828 (< 0.0001) (-0.082057, -0.059960)	-0.054776 (< 0.001) (-0.067369, -0.042943)	-0.026318 (< 0.001) (-0.038103, -0.017285)
Income ²	-0.055972 (< 0.0001) (-0.065602, -0.046343)	-0.041051 (< 0.001) (-0.051054, -0.032225)	-0.018364 (< 0.001) (-0.024970, -0.006941)
Income ³	-0.023682 (< 0.0001) (-0.031924, -0.015440)	-0.020436 (< 0.001) (-0.027514, -0.013662)	-0.005621 (< 0.001) (-0.011110, -0.002367)
Income ⁴	-0.013739 (0.0007) (-0.021703, -0.005776)	-0.008619 (0.008) (-0.014793, -0.002253)	0.000000 (0.544) (-0.004763, 0.002329)
Chronic conditions [†]	-0.036201 (< 0.0001) (-0.037904, -0.034498)	-0.030466 (< 0.001) (-0.033343, -0.027825)	-0.021045 (< 0.001) (-0.0232, -0.018970)

Note: First row in each cell is the coefficient estimate (*p*-value). The second row is the 95% confidence interval for the coefficient estimate.

*The first level of income adequacy, in comparison to the highest level of income adequacy.

†The number of self-reported chronic conditions.

symmetrically trimmed least squares model. Each of the three Tobit models found that gender was not significantly associated with health status ($p \geq 0.286$). Ordinary least squares estimation on the subjects with HUI scores below the ceiling showed that males had significantly lower health status than females ($p = .0196$). The coefficient estimates from the median regression model were identical to those from the censored LAD model.

The predictive accuracy of each model in an independent validation dataset is summarized in Table 3. The censored LAD and median regression models had the lowest median absolute prediction error (0.041). OLS regression on those subjects with

nonlimit HUI scores had the highest median absolute prediction error (0.057). The analysis was repeated separately among those aged at least 65 years, and among those younger than 50 years. In both sub-analyses, censored LAD and median regression had the lowest absolute prediction error.

Figure 2 describes the age-related decay in HRQL under each model, for a female in the highest income category with no chronic conditions. The classic Tobit model shows the steepest decline in HRQL with increasing age. The symmetrically trimmed least squares model shows moderate decline in HRQL until age 72. From age 72 to 80, the model describes HRQL as marginally

Table 2 Coefficient estimates for each of the three models that ignore censoring in the HUI score

Coefficient	OLS estimation on entire dataset	OLS estimation on subjects with HUI < 1	Median regression
Intercept	1.01976 (< 0.0001) (1.00070, 1.03881)	0.96776 (< 0.0001) (0.94491, 0.99061)	0.988816 (< 0.001) (0.953187, 1.008620)
Age	-0.00210 (< 0.0001) (-0.00290, -0.00130)	-0.00144 (0.0028) (-0.00238, -0.00050)	-0.001217 (0.086) (-0.001960, 0.000238)
Age ²	0.00001 (0.0065) (0.00000, 0.00002)	0.00001 (0.0131) (0, 0.00002)	0.000007 (0.320) (-0.000007, 0.000015)
Male	-0.00223 (0.2768) (-0.00624, 0.00179)	-0.00563 (0.0196) (-0.01035, -0.00090)	0.002636 (0.280) (-0.000760, 0.003727)
Income ^{1*}	-0.06135 (< 0.0001) (-0.07048, -0.05222)	-0.06599 (< 0.0001) (-0.07654, -0.05543)	-0.026318 (< 0.001) (-0.039302, -0.017118)
Income ²	-0.04953 (< 0.0001) (-0.05732, -0.04173)	-0.05586 (< 0.0001) (-0.06498, -0.04674)	-0.018364 (< 0.001) (-0.025169, -0.006667)
Income ³	-0.02234 (< 0.0001) (-0.02896, -0.01571)	-0.02838 (< 0.0001) (-0.03628, -0.02048)	-0.005621 (< 0.001) (-0.011400, -0.002333)
Income ⁴	-0.01243 (0.0001) (-0.01882, -0.00603)	-0.01505 (0.0001) (-0.02269, -0.00741)	0.000000 (0.558) (-0.004667, 0.002404)
Chronic conditions [†]	-0.03218 (< 0.0001) (-0.03358, -0.03078)	-0.03141 (< 0.0001) (-0.03297, -0.02986)	-0.021045 (< 0.001) (-0.023478, -0.019102)

Note: First row in each cell is the coefficient estimate (*p*-value). The second row is the 95% confidence interval for the coefficient estimate.

*The first level of income adequacy, in comparison to the highest level of income adequacy.

†The number of self-reported chronic conditions.

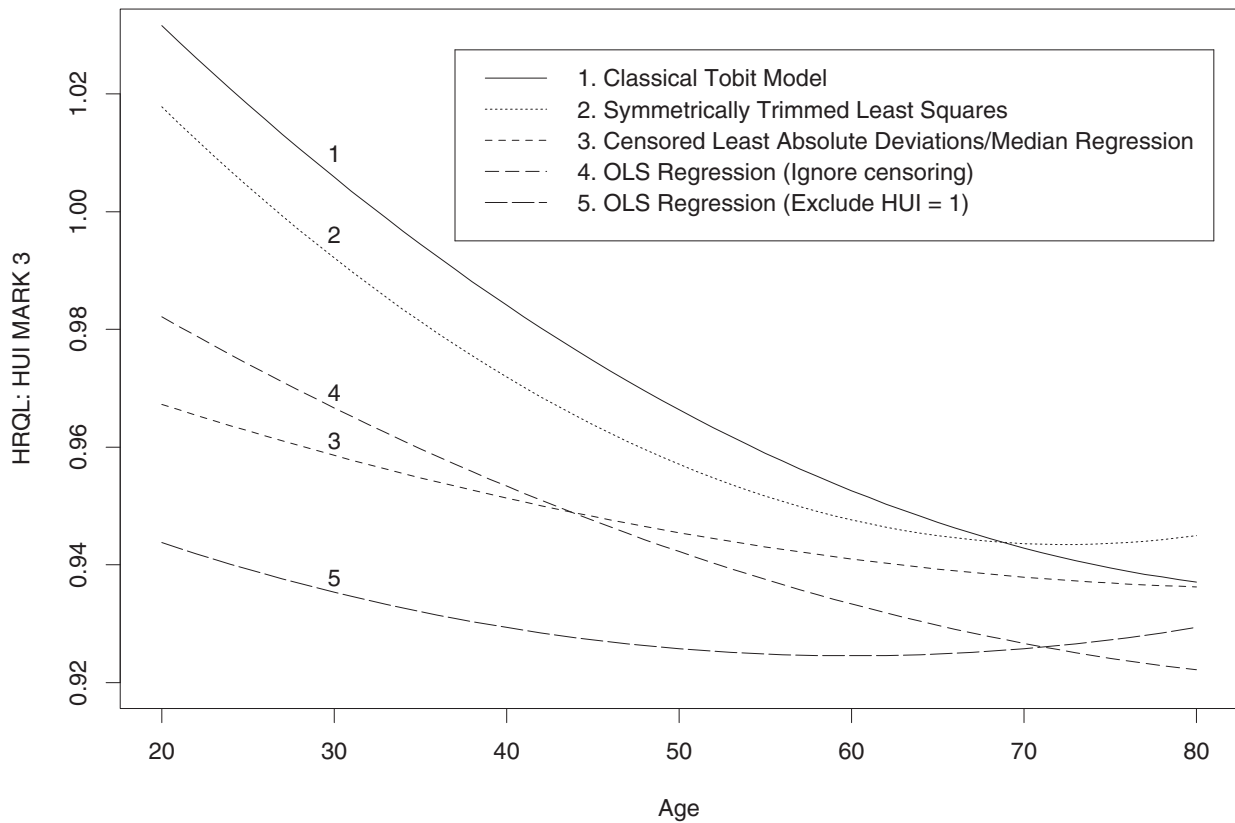


Figure 2 Age-related decay in health-related quality of life (female in the highest income category with no chronic conditions).

Table 3 Predictive accuracy of each model in an independent validation dataset

Model	Absolute prediction error*
Classical Tobit model	0.056 (0.026–0.104)
Symmetrically trimmed least squares	0.050 (0.023–0.099)
Censored LAD	0.041 (0.017–0.092)
OLS on all subjects	0.050 (0.021–0.095)
OLS on subjects with HUI < 1	0.057 (0.030–0.098)
Median regression	0.041 (0.017–0.092)

Absolute prediction error is summarized by the median (first quartile to third quartile).

improving with increasing age. The censored LAD model and the median regression model show a relatively modest decline in HRQL with increasing age. The OLS model showed a continuous decline in HRQL with increasing age. The OLS model fit to the nonlimit subjects showed the lowest average HRQL over most of the age range. It also showed that HRQL declined with increasing age, until approximately age 60 after which the model described HRQL as increasing with age to 80 years. Over the range in which HRQL decreased with increasing age, all models showed that the rate of

decay in HRQL decreased with increasing age. The same pattern would be observed for different combinations of subject characteristics with each of the six models, but the regression curve would be shifted upwards or downwards by a fixed amount.

Discussion

Measures of health status are frequently reported in population health surveys. Often, investigators wish to relate health status to characteristics such as age, gender and socioeconomic status. Measurements of health status can be subject to a ceiling effect. A perfect score can be interpreted as a censored observation in that the index is not sensitive enough to determine gradations in health status among those above a certain threshold. This paper compares the performance of the classic Tobit model estimated using maximum likelihood methods, with that of two robust alternatives to the Tobit model. For the purpose, of comparison three models that ignore the ceiling effect were also fit to the data.

There were several similarities in the results across models. Each model described health status

as increasing with increasing income adequacy, and with a decrease in number of reported chronic conditions. All three Tobit models found that gender was not associated with health status, and described HRQL as decreasing with increasing age at least until age 72. However, there were also important differences between the three models. The three Tobit models differed in the estimated magnitude of the effect of age on HRQL. The rate of decrease in HRQL with increasing age was much lower for the censored LAD model than for the other two models. The symmetrically trimmed least squares model shrunk the estimates of the effects of income adequacy and chronic conditions toward the null value compared to the classic Tobit model. The censored LAD model shrunk the effects for these variables toward the null, compared to the other two Tobit models. Of the six models, the censored LAD and median regression models had the lowest absolute prediction errors in an independent validation sample.

The coefficient estimates for the median regression model were the same as those for the censored LAD model. Since the fitted values for the median regression model were all less than 1, the iterative algorithm for estimating the censored LAD coefficients terminated after a single iteration. This suggests that censoring only affects the upper tail of the distribution of health status for all values of the regressor, and that the median is below the censoring percentile for all covariate patterns in the data. If the percentage of censored observations were to be in excess of 50%, then it is likely that the two models would produce different coefficient estimates.

The description of the effect of age on HRQL was potentially problematic for all three Tobit models. All three models described the rate of decay in HRQL as decreasing with increasing age at least until age 72. Intuitively, one would expect that the rate of decay in HRQL would either increase with increasing age, or at least remain constant. One possible reason for the counterintuitive effect of age is that the NPHS purposely excludes residents of long-term care facilities. This introduces the possibility of a healthy-survivor effect: the community dwelling elderly may be healthier on average than those marginally younger. Furthermore, it is possible that response rates were related to health status, with less healthy subjects being less willing to participate in the NPHS. Furthermore, it should be noted that the HUI is a preference measure, not a direct utility measure. It is possible that one adjusts one's sense of well-being to one's relative state of

health. People who are getting older expect to have limited health, and so minor disability doesn't really affect their sense of well-being.

There are certain limitations to our present study. This study was not intended to be an exhaustive study of the Health Utilities Index and its relationship to socio-demographic characteristics. We could have fit regression models with a larger number of predictor variables, or used models that examined the interactions between different variables. However, the purpose of the paper was to explore the performance of three different methods for estimating regression models in the presence of a ceiling effect. Therefore, we chose a small number of regressors that we felt, a priori, would be related to health status. Statisticians in medical research have developed flexible parametric models for fitting regression models to survival data [19]. These methods allow one to specify the distributional form of the response variable. However, for the current study we chose to limit our attention to robust econometric methods that have previously been proposed for analyzing HRQL. Similarly, the two-stage model [20] was not considered as an alternative, although it could be examined in future research. A further limitation to the analysis was that only the frequency but not the type of chronic medical condition was considered in the analysis. However, the purpose of the analysis was not an exhaustive examination of the effect of demographic variables on health status, but on how the results differed across models.

Accurate estimation of the relationship between health status and predictors of health is an important issue in population health research, and has implications for public health. The absolute differences between the coefficient estimates for different models are not large. However, the unit of the slope coefficient corresponds to the change in health status for a one-unit change in the predictor. In our data, the first and third quartiles of the HUI score were 0.877 and 0.947, respectively. Most observable HUI scores lie in a small range. Therefore, a small difference in an estimated parameter can have an important impact on our interpretation of the strength of a health indicator's effect on health status.

If HRQL scores are subject to a ceiling effect, ignoring the censoring that occurs at the ceiling can have negative consequences. Greene [21] demonstrated that ignoring the censoring and fitting regression models estimated using OLS results in coefficient estimates that are systematically biased toward 0. Furthermore, one cannot circumvent the

problem by restricting the analysis to those subjects whose HRQL scores lie below the ceiling value, since this too will result in biased coefficient estimates [21]. This is borne out in that coefficient estimates for the OLS model that ignores censoring tend to be smaller than those obtained from the classic Tobit model. Previous research has shown that it is likely that the conditional distribution of the HUI, conditional on similar regressors is heteroscedastic [10]. Hence, it is likely that the coefficient estimates for the Tobit model are themselves subject to bias. The relationship between the coefficients for the Tobit model and the symmetrically trimmed least squares model was inconsistent, making it difficult to assess the direction of the possible bias in the OLS and Tobit coefficients.

The Tobit model is a statistical tool that allows one to circumvent possible ceiling effects in measures of health status. However, the classic Tobit model, estimated using maximum likelihood methods, is susceptible to both non-normality and heteroscedasticity of the error term, both of which can result in biased coefficient estimates. Robust alternatives to maximum likelihood estimation have been proposed that relax the assumptions of normality and of uniform variance. We have compared the performance of three different estimation methods for the Tobit model. All three Tobit models illustrated similar trends across models for most of the predictor variables.

The methods examined in this study can also be used for studying HRQL scores that are subject to a floor effect, and not just to a ceiling effect. A floor effect may occur if someone with an observed HUI score of -0.03 (the lowest possible observed HUI score) had a true HUI less than -0.03 . In the NPHS data the lowest observed HUI score was 0.077 . Since no one in the NPHS had an observed HUI score of -0.03 , it was not necessary to incorporate the existence of a floor effect into the reported analyses. In clinical research, it is possible that many patients would take on the floor value of the HRQL score, and then the abilities of methods to incorporate floor effects would become important.

The results obtained in this study and the differences between models highlight the need for careful consideration of how to fit statistical models to measures of health status that may be subject to a ceiling effect. The current study examined robust econometric tools for fitting regression models in the presence of censored data. The results were compared to those from three models that ignored the censoring in the HUI score. We propose the use of the censored LAD model to analyze HRQL data

for two reasons. First, from a theoretical perspective, this model is robust to heteroscedasticity and non-normality of errors and does not require that the conditional distribution be symmetric and thus, has weaker assumptions than either the classic Tobit model or the symmetrically trimmed least squares model. Second, this model had the lowest prediction error in an independent validation dataset. Furthermore, this model produced similar results to the classic median regression model. Regardless of one's view of the likelihood of the presence of censoring in HRQL scores, one will obtain similar results using either method. Both models describe the relationship between median health status, rather than mean health status, and subject characteristics. Given the skewed distribution of HUI scores, the median may be a more valid indication of central tendency than the mean.

References

- 1 Stewart AL, Hays RD, Ware JE. Methods of constructing health measures. In: Stewart AL, Ware JE, eds. *Measuring Function and Well-Being. The Medical Outcomes Study Approach*. Durham, NC, and London: Duke University Press, 1992.
- 2 Froberg DG, Kane RL. Methodology for measuring health-state preferences—I: measurement strategies. *J Clin Epidemiol* 1989;42:345–54.
- 3 Feeny DH, Torrance GW, Furlong WJ. Health Utilities Index. In: Spilker B, ed. *Quality of Life and Pharmacoeconomics in Clinical Trials* (2nd ed.). Philadelphia: Lippincott-Raven Press, 1996; 239–52.
- 4 Austin PC, Escobar M, Kopec J. The use of the Tobit model for analyzing measures of health status. *Qual Life Res* 2000;9:901–10.
- 5 Grootendorst P. Censoring in statistical models of health status: what happens when one can better than '1'. *Qual Life Res* 2000;9:911–4.
- 6 Powell J. Symmetrically trimmed least squares estimation for the Tobit models. *Econometrica* 1986;54:1435–60.
- 7 Powell J. Least absolute deviations estimation for the censored regression model. *J Econometrics* 1984;25:303–25.
- 8 Maddala GS. *Limited-Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press, 1983; 178–82.
- 9 Greene WH. *Econometric Analysis* (4th ed.). Upper Saddle River, NJ: Prentice Hall, 2000.
- 10 Austin PC. Bayesian extensions of the Tobit model for analyzing measures of health status. *Med Decis Making* 2002;22:149–59.
- 11 Statistics Canada. *NPHS Public Use Microdata Documentation*. Ottawa, Canada: Statistics Canada, 1995.

- 12 Torrance GW, Feeny DH, Furlong WJ. Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. *Med Care* 1996;34:702–22.
- 13 Tobin J. Estimation of Relationships for Limited Dependent Variables. *Econometrica* 1958;26:24–36.
- 14 Johnston J, DiNardo J. *Econometric Methods* (4th ed.). New York: McGraw-Hill, 1997.
- 15 Paarsch H. A Monte Carlo Comparison of Estimators for Censored Regression Models. *J Econometrics* 1984;24:197–213.
- 16 Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1993.
- 17 SAS Institute Inc SAS, Version 8.2. Cary, NC: SAS Institute Inc, 2001.
- 18 Statistical Sciences, Inc. S-Plus 5.1 Release 1. Seattle: StatSci, a division of MathSoft, Inc., 1999.
- 19 Cox DR, Oakes D. *Analysis of Survival Data*. London: Chapman & Hall, 1984.
- 20 Mullahy J. Much ado about two. reconsidering retransformation and the two-part model in health econometrics. *J Health Econ* 1998;17:247–81.
- 21 Greene WH. On the asymptotic bias of the ordinary least squares estimator of the Tobit model. *Econometrica* 1981;49:505–13.