DISEASE MANAGEMENT AND LATENT CHOICES

By

SEAN MICHAEL MURPHY

A dissertation submitted in partial fulfillment
of the requirements for the
degree of

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY
School of Economic Sciences

August 2008

To the Faculty of Washington State University

The members of the Committee appointed to examine the dissertation of SEAN MICHAEL MURPHY find it satisfactory and recommend that it be accepted.

_____
Chair

_____

_____

ACKNOWLEDGEMENTS

I would first like to thank the members of my doctoral committee: Dr. Robert Rosenman, Dr. Daniel Friesner, and Dr. Jonathan Yoder. I really can not express enough gratitude for the excellent guidance and the amount of support they have provided me with throughout my career as a graduate student at Washington State University.

I also want to thank my family for always supporting my decisions and providing me with the foundation to achieve my dreams. I would especially like to thank my mother, whose infinite optimism and pride have always inspired me to reach for the top.

Most of all, I want to thank my beautiful wife Antoinette. She is the love of my life, and without her unwavering love and support, none of this would have been possible. We have already shared a great number of adventures, and I can not wait to begin this next journey with her. It is to her that I dedicate this dissertation.

Sean M. Murphy

DISEASE MANAGEMENT AND LATENT CHOICES

Abstract

By Sean Michael Murphy, Ph.D.
Washington State University
August, 2008

Chair: Robert Rosenman

This dissertation consists of three independent essays in the field of health economics. The first essay analyzes factors influencing the initial treatment choice of inpatients with severe hypertension. A thorough analysis of treatment choice has been largely overlooked in the hypertension literature, and few studies of any disease have conducted comprehensive multivariate analyses on treatment choices using such a diverse array of socioeconomic variables and hospital locations. According to the results, characteristics other than morbidity affect the type of treatment received; indicating public policy could improve care.

The second essay analyzes the effect that a patient's reference point has on her perceived effectiveness of subsequent treatment. One commonly used measure of treatment effectiveness for conditions where treatments are palliative, and clear objective symptoms do not exist, is self-reported changes in disease status. Factors such as treatment history provide a reference point that may influence patients' expectations of how effective further treatment might be. Therefore, decisions about whether to proceed with additional treatment, as well as perceptions of how effective that treatment is, may

be influenced by this point. Although there is an extensive literature on how patient expectations influence treatment outcomes, work testing how expectations depend on these reference point factors appears to be missing. The results indicate that these factors influence perceived treatment effectiveness.

The final essay focuses on missing and ambiguous observations in a dataset with binary dependent variables. The ability to reallocate these responses could aid in the correction of potentially biased estimates. Using the "latent-choice multinomial logit model" (LCMNL), it is possible to determine whether these incomplete responses are more likely to belong to another outcome. Simulations of this model are performed to determine whether the estimated conditional probabilities are accurate enough to evaluate the likelihood that any given observation belongs to a particular outcome, and whether doing so improves parametric estimation. Tests imply that the reclassifications indicated by the LCMNL's conditional probabilities are accurate. The best method for dealing with ambiguous observations in empirical analysis is also assessed. Results indicate that the best method depends on the source of the ambiguity.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

The following are three independent essays in the field of health economics. All address issues concerning disease management, in that they either help develop a better understanding of how treatment decisions are made, along with the effects of those treatments, or focus on methods that help increase the accuracy of our analysis. Addressing these issues should help increase efficiency in the production of health.

The first essay analyzes the effects that patient's socioeconomic characteristics, along with hospital size and location, have on the initial treatment choice for inpatients in the Clinical Classification Software's (CCS) category 99, "Hypertension with Complications and Secondary Hypertension" (HCUP, 2000-2003). According to Singh (2006), hypertension is quite prevalent in that it affects approximately 25% of those 18 and over. However, many are unaware of the condition, and as a result do not seek treatment until complications have arisen. After being hospitalized, physicians and patients must choose from an array of initial therapeutic procedures, all of which can be categorized as either invasive or non-invasive. Invasive procedures are significantly more expensive than non-invasive, which implies that factors other than a patient's clinical condition may come into play when determining a treatment path.

This study makes several contributions to the treatment choice literature. Until now, a thorough analysis of treatment choice has been largely overlooked in the hypertension literature; although, many of these socioeconomic factors have been considered in the treatment choice of other conditions. However, few studies of any disease have been able to control for such a diverse array of socioeconomic variables and hospital locations.

The second essay investigates the effect that a patient's reference point has on her perceived effectiveness of a given treatment. For some illnesses or health disorders, treatments are palliative rather than curative, and clear objective symptoms do not exist. One commonly used measure of treatment effectiveness in these instances is self-reported changes in disease status. We argue that factors such as treatment history provide patients with a reference point which influences their expectations of treatment success; this in turn affects not only their decision of whether additional treatment is pursued, but also their perceptions of how effective the treatment is should they decide to proceed. There is a fair amount of literature regarding the relationship between expectations, primarily optimism, and recovery speed (Frey et al., 1985; Kalauokalani et al., 2001; Scheier and Carver, 1987; Scheier et al., 1989), as well as subjective health (Carver et al., 1994; Koller et al., 2000; Llewellyn-Thomas, Thiel, and McGreal, 1992); however, these papers fail to address the determinants of these expectations.

Through the use of an adapted prospect theory application we model the effect a patient's reference point has on her expectations of treatment success, thus determining the path of treatment, and on the perceived success of treatments after they have been pursued. We expect that patients with unsuccessful prior treatments have a frame of reference leaving them less likely to expect improvement from subsequent treatments. Using a data set on Idiopathic Intracranial Hypertension we test the hypothesis implied by our model.

The final essay of this dissertation focuses on missing and ambiguous observations in a dataset that includes binary response dependent variables. This is a major concern for applied researchers of many fields, as estimation problems may arise due to this deficiency of information, leading to inaccurate results, and possibly limiting the types of analyses available to the researcher. The question of how best to treat the *don't know* (DK) response in contingent

valuation data has received quite a bit of attention (Alberini, Boyle, and Welsh, 2003; Carson et al., 1998; Groothuis and Whitehead, 2002; Haener and Adamowicz, 1998; Wang, 1997). Previous methods of dealing with these responses include: grouping the responses with the *no* category (Carson et al., 1998), excluding the responses (Johannesson et al., 1993; Wang, 1997), and treating them as a middle response between *yes* and *no* in ordered categorical models (Groothuis and Whitehead, 2002; Wang, 1997). Based on the assumption that some ambiguous responses are misclassified, Caudill (2006) extends Dempster, Laird, and Rubin's (1977) work, developing a multinomial logit model with missing information, titled the "latent-choice multinomial logit model" (LCMNL). Caudill and Groothuis (2005) use the model to statistically determine whether DK responses in the contingent valuation literature are actually more like a *yes*, a *no*, or truly a DK.

The purpose of this paper is to perform some simulations of the LCMNL to determine: 1) if the conditional probabilities generated by the model are in fact accurate enough to assess the likelihood that individual observations belong to a particular outcome as Caudill, Ayuso, and Guillen (2005) and Caudill and Groothuis (2005) did, and 2) the best course of action for dealing with incomplete responses in empirical analysis. Through the use of a specified data generating process, a random data matrix is created, intervariable correlations from a real world dataset imposed on it, and a dependent variable generated using known parameter values. As a result, the true values of the incomplete responses are known beforehand, which allows us to accomplish our first goal. Knowledge of the true parameter values allows us to accomplish the second objective by being able to test the coefficient estimates against their true values.

References

Alberini, A., K. Boyle, and M. Welsh. "Analysis of Contingent Valuation Data with Multiple Bids and Response Options Allowing Respondents to Express Uncertainty." *Journal of Environmental Economics and Management* 45, no. 1 (2003): 40-62.

Carver, C. S., C. Pozo-Kaderman, S. D. Harris, V. Noriega, M. F. Scheier, D. S. Robinson, A. S. Ketcham, F. L. Moffat, Jr., and K. C. Clark. "Optimism Versus Pessimism Predicts the Quality of Women's Adjustment to Early Stage Breast Cancer." *Cancer* 73, no. 4 (1994): 1213-20.

Carson, R. T., W. M. Hanemann, R. J. Kopp, J. A. Krosnick, R. C. Mitchell, S. Presser, P. A. Ruud, V. K. Smith, M. Conaway, and K. Martin. "Referendum Design and Contingent Valuation: The Noaa Panel's No-Vote Recommendation (Vol 80, Pg 335, 1998)." *Review of Economics and Statistics* 80, no. 3 (1998): 484-87.

Caudill, S. B. "A Logit Model with Missing Information Illustrated by Testing for Hidden Unemployment in Transition Economies." *Oxford Bulletin of Economics and Statistics* 68, no. 5 (2006): 665 - 77.

Caudill, S. B., M. Ayuso, and M. Guillen. "Fraud Detection Using a Multinomial Logit Model with Missing Information." *Journal of Risk and Insurance* 72, no. 4 (2005): 539-50.

Caudill, S. B., and P. A. Groothuis. "Modeling Hidden Alternatives in Random Utility Models: An Application To "Don't Know" Responses in Contingent Valuation." *Land Economics* 81, no. 3 (2005): 445-54.

Dempster, A. P., N. M. Laird, and D. B. Rubin. "Maximum Likelihood Estimation from Incomplete Data Via the Em Algorithm." *Journal of the Royal Statistical Society, Series B* 39 (1977): 1 - 38.

Frey, D., O. Rogner, M. Schuler, C. Korte, and D. Havemann. "Psychological Determinants in the Convalescence of Accident Patients." *Basic and Applied Social Psychology* 6, no. 4 (1985): 317-28.

Groothuis, P. A., and J. C. Whitehead. "Does Don't Know Mean No? Analysis of 'Don't Know' Responses in Dichotomous Choice Contingent Valuation Questions." *Applied Economics* 34, no. 15 (2002): 1935-40.

Haener, M. K., and W. L. Adamowicz. "Analysis of 'Don't Know' Responses in Referendum Contingent Valuation Questions." *Agricultural and Resource Economics Review* 27, no. 2 (1998): 219 - 30.

HCUP. "Clinical Classification Software (CCS) for ICD-9-CM." Downloaded from http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp, 04 August 2006.

Johannesson, M., P. O. Johansson, B. Kristrom, and U. G. Gerdtham. "Willingness to Pay for Antihypertensive Therapy--Further Results." *J Health Econ* 12, no. 1 (1993): 95-108.

Kalauokalani, D., D. C. Cherkin, K. J. Sherman, T. D. Koepsell, and R. A. Deyo. "Lessons from a Trial of Acupuncture and Massage for Low Back Pain: Patient Expectations and Treatment Effects." *Spine* 26, no. 13 (2001): 1418-24.

Koller, M., W. Lorenz, K. Wagner, A. Keil, D. Trott, R. Engenhart-Cabillic, and C. Nies. "Expectations and Quality of Life of Cancer Patients Undergoing Radiotherapy." *J R Soc Med* 93, no. 12 (2000): 621-8.

Llewellyn-Thomas, Hilary A., Elaine C. Thiel, and M. June McGreal. "Cancer Patients' Evaluations of Their Current Health States: The Influences of Expectations, Comparisons, Actual Health Status, and Mood." *Medical Decision Making* 12 (1992): 115-22.

Scheier, M. F., and C. S. Carver. "Dispositional Optimism and Physical Well-Being: The Influence of Generalized Outcome Expectancies on Health." *J Pers* 55, no. 2 (1987): 169-210.

Scheier, M. F., K. A. Matthews, J. F. Owens, G. J. Magovern, Sr., R. C. Lefebvre, R. A. Abbott, and C. S. Carver. "Dispositional Optimism and Recovery from Coronary Artery Bypass Surgery: The Beneficial Effects on Physical and Psychological Well-Being." *J Pers Soc Psychol* 57, no. 6 (1989): 1024-40.

Singh, V. N. “High Blood Pressure.” Downloaded from www.emedicinehealth.com, 04 August 2006.

Wang, H. "Treatment of "Don't-Know" Responses in Contingent Valuation Surveys: A Random Valuation Model." *Journal of Environmental Economics and Management* 32, no. 2 (1997): 219-32.

CHAPTER TWO

DETERMINING FACTORS IN THE TREATMENT CHOICE OF PATIENTS WITH HYPERTENSION WITH COMPLICATIONS AND SECONDARY HYPERTENSION

Abstract

We analyzed the effect patients' socioeconomic characteristics, along with hospital size and location, had on the initial treatment choice for inpatients with severe hypertension. A thorough analysis of treatment choice has been largely overlooked in the hypertension literature, and few studies of any disease have conducted comprehensive multivariate analyses on treatment choices using such a diverse array of socioeconomic variables and hospital locations. Even after accounting for the U.S. public health insurance programs it appears there are still uninsured individuals seeking inpatient care, and receiving potentially less effective treatments. Patients covered by Medicare are much more likely to receive the relatively expensive treatments, even after controlling for age and comorbidity, implying that Medicare patients are for some reason treated differently. Also the study shows that racial disparity and geographical treatment variation remain issues when it comes to treatment decisions. Finding that characteristics, other than morbidity, affect the type of treatment received indicates public policy could improve care.

Keywords: Hypertension, Treatment choice, Invasive, Geography, Logistic regression, Health insurance.

Introduction

Hypertension refers to the sustained condition of abnormally high blood pressure. Left untreated, hypertension can lead to serious complications such as heart or kidney failure. Approximately 25% of people 18 and over have high blood pressure; however, many are unaware of the condition (Singh, 2006). As a result, treatment is often not sought until complications arise. Once a patient is hospitalized, physicians and patients must choose from an array of initial therapeutic procedures, all of which can be categorized as either invasive or non-invasive. Considering that the costs of the invasive procedures are significantly higher, decisions about what treatment path to follow may depend on more than just the patient's clinical condition, (average total charges for patients prescribed invasive procedures exceeded those for patients with non-invasive procedures by about $9000). Whether ability to pay, insurance coverage or geographic location influence care path is our primary interest. It is also possible that demographic variables such as age, sex and race influence the initial prescribed treatments, although differences correlated with these variables *may* be from different effectiveness of non-invasive and invasive procedures according to these characteristics. To explore this topic, we used a binary logistic regression on treatment choice to test whether a patient's socioeconomic characteristics, along with hospital size and location, influenced the initial treatment choice for individuals in the Clinical Classification Software's (HCUP, 2000-2003) category 99 "Hypertension with Complications and Secondary Hypertension".

Our study makes several contributions to the treatment choice literature. First, as opposed to previous studies focusing on geographical regions generally no larger than the state level, we use inpatient data from hospitals in 25 states. We are able to control for a relatively large number of socioeconomic variables. This is something that seems to have been largely overlooked in

hypertension treatment, although socioeconomic factors have been considered in the treatment choice for different types of cancer and heart disease (Bergman *et al.,* 1991; Desch *et al.,* 1996; Greenberg *et al.,* 1988; Bradley, Given, and Roberts*,* 2002; Roetzheim *et al.,* 2000; Samet *et al.,* 1986; Satariano, Swanson, and Moll*,* 1992; Wenneker and Epstein, 1989; Whittle *et al.,* 1993), as well as depression (Sturm, Meredith, and Wells*,* 1996). In the next section, we explore the theory behind non-clinical treatment variation and provide a brief literature review. Then, we describe the data before explaining the model we used in our study. Next, we present the results of our estimation while the last section discusses the implications of our analysis and suggests future research.

Factors influencing treatment choice

Although many studies have used retrospective data analysis to examine the effects of socioeconomic variables on treatment choice, little has been done on the effect of such variables on hypertension, and few treatment choice studies of any disease have been able to control for such a wide range of explanatory variables. Many of these studies have looked at various types of cancers. Among the most common socioeconomic predictors is age. Greenberg *et al.* (1988) found that older patients were less likely to receive surgery versus no surgery, and other forms of treatment versus no treatment for patients diagnosed with lung cancer. Desch *et al.* (1996) also found that increasing age reduced the likelihood of treatment versus no treatment, and surgery versus radiation in patients with prostate cancer. Age was also documented by Bergman and others (1991); while studying patients 55 years and older with breast cancer, the authors found that patients 75 years and older were less likely to receive radiation and more likely to be given hormonal therapy. Common reasons given for the age effect are frailty and co-morbidity (Samet

et al., 1986). Both Greenberg *et al.* (1988) and Desch *et al.* (1996) found this effect to persist even after controlling co-morbidity, which was statistically significant only in the latter study. One study found that older women were significantly more likely to receive the relatively invasive procedure for treatment of breast cancer (Satariano, Swanson, and Moll*,* 1992).

The effect of a patient's sex on treatment choice is not well documented. Studies that took gender into account primarily deal with treatment choices for different cancers. Many studies review treatments for diseases that are gender specific (Bergman *et al.,* 1991; Desch *et al.,* 1996; Roetzheim *et al.,* 2000; Satariano, Swanson, and Moll*,* 1992). Others found it insignificant (Desch *et al.,* 1996; Greenberg *et al.,* 1988). The literature pertaining to race's effect on treatment choice is vast. Most findings point to racial disparity between white and black patients. Desch *et al.* (1996) indicated that the odds of white patients with prostate cancer receiving surgery were approximately three times greater than the odds of black patients after controlling age, sex, income, co-morbidity, residence (rural versus urban) and education. Wenneker and Epstein (1989) and Whittle *et al.* (1993) found white patients remained significantly more likely to undergo invasive cardiac procedures even after accounting age, sex, income, co-morbidity, and payer type (Wenneker and Epstein, 1989; Whittle *et al.,* 1993) or physician financial incentives (Whittle *et al.,* 1993). Sturm, Meredith, and Wells (1996) reported white patients were more likely to seek specialty psychiatric care, i.e., "psychiatrist" versus "general medical sector," and "non-physician mental-health specialist" versus the "general medical sector," relative to non-white patients. The authors also controlled for health status, age, sex, education, income and payment method.

However, the evidence is mixed on the effect race has on women's breast cancer treatment choice. Bradley, Given, and Roberts (2002) analyzed women who received either no

surgery, breast conserving surgery, breast conserving surgery with radiation or mastectomy. The authors found that African-American women were less likely than white women to undergo surgery and out of those having surgery, African-American women were more likely to have breast conserving surgery than white women. Other covariates in the model included age, disease severity, income and whether the patient was covered by Medicaid. The only racial difference discovered by Roetzheim *et al.* (2000) was that Hispanic patients were more likely to receive breast conserving surgery than non-Hispanic whites. Like many of the other studies mentioned above, the authors controlled age, education, residence, income, co-morbidity and payer type. Satariano, Swanson, and Moll (1992) found race to be insignificant after accounting for variables such as age, sex and hospital size. Income, health status and payment method were not included in their stepwise logistic regression. Income and payment method are often found to be important treatment choice predictors. This is usually attributed to certain treatments being more expensive than others. In several studies, income and the probability of receiving the relatively expensive treatment were positively correlated (Bradley, Given, and Roberts, 2002; Desch *et al.,* 1996; Sturm, Meredith, and Wells, 1996) although at least one paper found that income did not play a significant role in the type of therapy chosen (Roetzheim *et al.,* 2000). Greenberg *et al.* (1988) found that lung cancer patients with private insurance were more likely to be surgically treated, or if surgery was not used then these individuals were more likely to receive another form of treatment than individuals with other forms of insurance, or none at all.

Sturm, Meredith, and Wells (1996) observed that being in a prepaid plan, relative to a fee-for-service plan, significantly reduced the likelihood that depressed patients would see a psychiatrist and increased the likelihood that these patients would see a non-physician mental-health specialist. This was anticipated since specialists are usually expensive relative to general

medical practitioners and cost containment would be expected to play a larger role in prepaid plans. Medicare HMO patients were more likely to receive breast conservation surgery than patients with fee-for-service Medicare (Roetzheim *et al.,* 2000). And among non-Medicare patients, those with no insurance were less likely to receive the same surgery than those with commercial fee-for-service insurance plans. Bradley, Given, and Roberts (2002) found individuals covered by a fee-for-service Medicaid plan were less likely to receive breast conserving surgery and more likely to receive no surgery than patients were not covered by Medicaid.

Little work has explored the influence of hospital size on treatment choice. Satariano, Swanson, and Moll (1992) found that women with early-stage breast cancer, treated in larger hospitals, were more likely to receive relatively less-invasive procedures than those treated in smaller hospitals. The authors attribute this finding to the possibility that large hospitals may be able to implement new forms of therapy more quickly. Desch *et al.* (1996) also found that patients residing in an urban area were more likely to receive some form of treatment. The patient's residence may serve as a proxy for hospital size since urban areas generally have larger hospitals. Urban residents also appear to be more likely to have breast conserving surgery (Roetzheim *et al.,* 2000).

The existence of medical practice variations across various geographic areas is a topic that has received a great deal of attention (Phelps, 1992). Studies compared US regions to other countries (McPherson *et al.,* 1981; McPherson *et al.,* 1982), differences within a specific state (Lewis, 1969; Wennberg and Gittelsohn, 1973; Wennberg and Gittelsohn, 1975; Phelps and Parente, 1990), among states (Ahronheim *et al.,* 2001; Krumholz *et al.,* 1998), US census regions (Pilote *et al.,* 1995) and among regional health care markets (O'Connor *et al.,* 1999). Specific to

hypertension, Reynolds *et al.* (2003) analyzed geographic treatment variations in China. However, the analysis of geographic treatment variation in the US appears to be nonexistent.

Most treatment variation studies have significant limitations. It is common for a study to focus on a relatively small geographic region, usually no larger than the state level (Bradley, Given, and Roberts, 2002; Desch *et al.,* 1996; Greenberg *et al.,* 1988; Roetzheim *et al.,* 2000; Samet *et al.,* 1986; Satariano, Swanson, and Moll, 1992; Sturm, Meredith, and Wells, 1996; Wenneker and Epstein, 1989). Several control for sex by analyzing gender specific diseases such as prostate and breast cancer (Bergman *et al.,* 1991; Bradley, Given, and Roberts, 2002; Desch *et al.,* 1996). And some are not able to control a wide range of socioeconomic characteristics (Bergman *et al.,* 1991; Satariano, Swanson, and Moll, 1992). By contrast, we analyzes patients with a gender-neutral disease from a sample including data from hospitals in 25 states. In addition to the patient gender, we include a wide array of socioeconomic characteristics, such as: age; race; income; payment type; co-morbidity; admittance status and we are able to control for hospital size and location, allowing us to test whether a variety of socioeconomic characteristics affect treatment choice.

Data

Data for this study came from the Healthcare Cost and Utilization Project (HCUP) Nationwide Inpatient Sample (NIS) for 2003, a stratified probability sample of hospital inpatients. Strata were defined based on: (i) geographic region; (ii) management (public, voluntary or proprietary); (iii) location; (iv) teaching status and (v) bed size. In order for a hospital to be included in the sample, it must be a community, non-rehabilitation hospital, contained in the State Inpatient Database (SID) and match the corresponding American Hospital Association (AHA) Annual

Survey data (HCUP, 2006). In 2003, 3763 hospitals from 37 geographically dispersed states were included in the sample. The NIS attempts to represent the population of all acute care discharges from US hospitals of the type just described. Up to 20% of the hospitals in each stratum were randomly selected, producing a sample of 994 hospitals - 21% of 4,836 hospitals in the target population.

Our study focuses on patients with an initial diagnosis of hypertension with complications and secondary hypertension. Individuals were grouped into this category by the Clinical Classification Software (HCUP, 2000-2003). The CCS is a classification system developed at the Agency for Healthcare Research and Quality (AHRQ). It groups International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnoses and procedures into meaningful categories. After eliminating observations with missing values, we were left with 17,437 observations from 25 geographically dispersed states. Since the base sample is designed to represent the population, as mentioned in the paragraph above, our final sample should still reflect that after controlling for factors that may bias us away from the representative sample. To check this, we randomly selected 500 observations from the dataset composed of all individuals in CCS 99. A dummy variable was created indicating whether or not an observation contained missing information and would therefore be dropped. We then tested for significant differences across this variable using ANOVA and chi-square tests. This process was repeated a handful of times, and there were only two variables with clear differences. The average age appears to be significantly higher among the dropped observations and it seems that there are a significantly larger number of individuals among the dropped observations who were admitted electively. Table 1 defines our variables. We divided the primary diagnostic and therapeutic procedures performed into two categories, “invasive”, e.g. angioplasty, and “non-

invasive", e.g. injection of an anticoagulant, according to definitions describing ICD-9-CM codes. All procedures can be classified as one or the other. We note that invasive procedures are generally more expensive than those that are non-invasive. As a result of the cost difference, the initial treatment decision made by the physician and patient may vary based on socio-economic characteristics in addition to those that are health related.

The dependent variable, "invasive," takes the value of one if the initial procedure performed on the patient was invasive and zero if it was non-invasive. The "age" variable measures the age of the patient at the time of admission, while "age2" represents the age variable squared. The age2 variable allows us to test whether the marginal effect that age has on the probability of individuals receiving invasive procedures differs between younger and older patients. "Female" is a dummy variable indicating patient' sex. "Middle," "uppermid," and "upper" are binary income variables representing three of the income quartiles classified by the NIS. The income quartiles are defined by the median income of the zip code where the patient resides. Therefore, these variables are not a true measure of the household's income, but really measure the affluence of the patient's locality, which will depend, among other things, on whether the patient resides in an urban or rural area and the country area. "Lower" is the excluded category. Race variables indicate whether the patient is "Black," "Hispanic," "Asian" (which includes Pacific Islander), or "otherace". Other race does not include whites, which is the excluded category. "Selfpay," "Medicare," "Medicaid," "nocharge," and "otherpay" are all dummy variables indicating the expected primary payer of the patient's medical bill. If the expected primary payer was something or someone other than private insurance, self pay, Medicare or Medicaid then the patient was assigned to the "otherpay" category. Thus, private insurance is the excluded category for the primary payer. "Elective" is a dummy variable that

equals one if the patient was admitted to the hospital on his or her own accord. We accounted for co-morbidity with the "ndx" variable - simply the total number of diagnoses on the patient's discharge record. "Bedsize," categorizes the hospital based on the number of beds and serves as a proxy for hospital size. Larger hospitals are often associated with medical centers, research programs and physician training (Satariano, Swanson, and Moll, 1992), and as a result may be more likely to have access to the most up-to-date treatment methods. Finally, 24 dummy variables were included to test for state-specific variations in treatment, with Florida serving as the base case.

Summary statistics for the entire sample and separated for those who did and did not receive invasive procedures are given in Table 2. Approximately 79% of the entire sample received some type of invasive procedure as the initial treatment for their hypertension. To explore the differences between patients in these categories, chi-square tests based on cross-tabulations were performed on the categorical variables. The Kruskal-Wallis rank test was run on the continuous variables. These results are presented in Table 3. For simplicity, all Kruskal-Wallis test results are placed in brackets and listed at the top of the table. Chi-square tests are presented below the Kruskal-Wallis tests. When conducting these tests, we did not correct for type I error, despite the fact that the sample is relatively large and is drawn from a finite population. Our rationale for this decision was based on two facts 1) that we do not know the true value of the population, thus any correction would be subject to error based on our approximation of the population size, and 2) that the sample size, while large, is nonetheless very small (less than ten percent and arguably close to five percent) compared to the size of the population. As such, while it is possible to implement an approximate finite population correction factor to ensure an appropriate level of significance and adjust for Type I error, doing

so adds little to the results of the tests. After any reasonable approximation correction our inferences remain the same**.** Also, degrees of freedom are not listed due to the large number, which implies that the t-statistics and z-statistics converge.

We found significant differences in the socio-economic characteristics of the two subsamples. Overall, our sample appears to be consistent with the hypertension prevalence rates for US adults (Glover *et al*., 2005). Given that the NIS is designed to be representative of the underlying population, these results are not surprising.

The patients' ages ranged from 1 to 103 years, with a mean of approximately 61 years. Although the range is large, the standard deviation is relatively small, giving a coefficient of variation (a normalized measure of variation, calculated by the ratio of the standard deviation to the mean) of only 0.28, hence most of our sample is concentrated around the mean. The prevalence rate for US adults 60 years of age or older, is approximately 65.2% (Glover *et al*., 2005). The sub-sample patients' average were 60 years (invasive) and 66 (non-invasive). The t-statistic testing for differences in the average age of these subpopulations is 18.63 ($p<.0001$), indicating a clear and significant variation. This finding is corroborated in Table 3; where the Kruskal-Wallis test indicates average age differed between those who had invasive procedures as their primary treatment and those who had non-invasive procedures. The sub-samples also differed by gender and race. Approximately half the patients in the entire sample and the invasive sample were female. This would also appear to coincide with the US male and female prevalence rates in that both populations are very near 30% (Glover *et al*., 2005). Of those receiving non-invasive procedures, 58% were female, which, as demonstrated in Table 3, is a significantly higher percentage than the sample as a whole. Taken in tandem, these statistics imply that the entire sample may slightly (by approximately 1-2%) over-represent the prevalence

of hypertension in females. As such, we must account for this possibility by controlling for gender in our empirical analysis. Overall approximately 43% of patients were white, 40% were Black, 12% were Hispanic, 2% were Asian or Pacific Islander and 3% were other races, indicating the disease is disproportionately prevalent among Blacks. The prevalence rates for US adults are approximately 27% White, non-Hispanics; 40.5% for Black, non-Hispanics; and 25.1% for Mexican Americans (Glover *et al*., 2005). As indicated in Table 3, the sample of patients receiving invasive procedures had a statistically significant greater share of Blacks and Hispanics compared to the sample receiving non-invasive treatments. There was no statistically significant difference in the percentages of Asians and other races, indicating that the invasive group must have a smaller share of whites.

Roughly 25% of the patients in the three samples were listed as residing in a zip code where the median income was in the range of $36,000 - $44,999. Twenty-percent were from a zip code with a median income in the $45,000 - $59,999 and 15% were listed as having come from a zip code with a median income in excess of $60,000. These also differed significantly between sub-samples, with a greater share of those receiving invasive treatments coming from the middle income group, while a greater portion of those receiving non-invasive treatments were from upper middle or upper economic zip codes. Interpolation tells us that a lower share of those receiving non-invasive treatments come from lower income zip codes.

Insurance coverage also differed across sub-samples. Approximately 63% of the entire sampled individuals were covered by Medicare, 18% by private insurance, 13% by Medicaid, and 2% were paying the bill by some other means. However, as indicated in Table 3, there was a statistically significant difference in the share of those receiving invasive treatments that were covered by Medicare (larger) and the share that were self-pay (smaller) when compared to those

receiving non-invasive treatments. There was no statistically significant difference in Medicaid and other pay between the two groups, with chi-square test statistics of .08 (p=.7733) and 1.32 (p=.2499), respectively.

Eighteen percent of the individuals in the study were admitted electively, with a much greater share of those receiving invasive procedures being elective admissions. The typical patient had approximately 7.5 diagnoses on his or her discharge record. The mean number of confounding diagnoses for those receiving invasive treatments was 7.4 compared to 7.7 for those receiving non-invasive treatment. This difference was statistically significant with a t-statistic of 6.25 (p<.0001). Those receiving invasive procedures were more likely to be in larger hospitals and to incur greater total charges when compared to those receiving non-invasive procedures.

With the exception of Texas, Florida, and New York, which accounted for approximately 16%, 14% and 13% of the patients in the entire sample respectively, individuals were distributed fairly evenly among the included states. Either the hypertension incidence is greater in Texas, Florida and New York or these states are over-represented in the sample (proportions were similar in the two sub-samples). In every state except Connecticut, Hawaii, Kansas, Maryland, North Carolina and Virginia, a significantly higher percentage of patients received invasive procedures as the initial form of treatment according to Table 3 (the results of this test for Vermont are questionable as a result of having too few observations in the two treatment groups). These findings imply that it is necessary to control for geography in our empirical analysis in order to ensure that our results can be generalized to make inferences about the population as a whole.

Method

We followed the retrospective data analysis approach similar to the studies mentioned above. However, we were able to control a wider range of socioeconomic characteristics than most of the previous articles. For example, we examined a relatively gender neutral disease and was able to control patient sex, allowing us to study its influence on treatment choice. We were also able to control the location of the hospital by state. When dealing with a dependent variable that is qualitative, researchers often use either a logistic or a probit model. In practice, the choice between logit and probit models is generally one of convenience (Gujarati, 1995). We chose a logistic model as it allows odds ratios for independent variables to be calculated. In a logistic regression, the dependent variable is transformed into a logit, which is the natural log of the odds. After this transformation, maximum likelihood estimation is applied to the model. The independent variable coefficients can be interpreted as the effect that a one unit change in the independent variable will have on the log odds. The odds ratios are useful when interpreting the parameter estimates of binary independent variables. They allow one to compare the odds that the dummy variable equaling one will lead to the dependent variable achieving a larger value relative to the odds for the base case. We performed a binary logistic regression in our study. Consistent with our previous discussion, the following model was used:

$$\begin{aligned} Invasive = {} & \beta_0 + \beta_1 Age + \beta_2 Age2 + \beta_3 Female + \beta_4 Black + \beta_5 Hispanic + \beta_6 Asian + \beta_7 Otherace \\ & + \beta_8 Middle + \beta_9 Uppermid + \beta_{10} Upper + \beta_{11} Selfpay + \beta_{12} Medicare + \beta_{13} Medicaid \\ & + \beta_{14} Otherpay + \beta_{15} Elective + \beta_{16} Ndx + \beta_{17} Bedsize + \beta_{18} AZ + \beta_{19} CO + \beta_{20} CT \\ & + \beta_{21} HI + \beta_{22} IA + \beta_{23} IN + \beta_{24} KS + \beta_{25} MA + \beta_{26} MD + \beta_{27} MI + \beta_{28} MO + \beta_{29} NC \\ & + \beta_{30} NH + \beta_{31} NJ + \beta_{32} NY + \beta_{33} PA + \beta_{34} RI + \beta_{35} SC + \beta_{36} TN + \beta_{37} TX + \beta_{38} UT \\ & + \beta_{39} VA + \beta_{40} VT + \beta_{41} WI + \varepsilon. \end{aligned}$$

The model's significance was tested using a Likelihood Ratio test. This test can be used much like the F-test in the classical linear regression model where the null hypothesis that all

coefficients in the model are equal to zero is tested. The model's likelihood ratio is used to generate a test statistic with a chi-square distribution. A sufficiently large test statistic implies a rejection of the null hypothesis. A likelihood ratio test was also run on the state (i.e., geographical) variables to establish whether they should be grouped into regions, or included separately. States were categorized into regions as defined by the US Census Bureau. The test was highly significant, indicating that state level control is appropriate.

Owing to the high age variable mean, a test was run to determine if its effect on the dependent variable would be better explained with a quadratic function. A Wald test was used to assess whether or not age and age2 were jointly significant. The Wald test can be used much like the F-test in the classical linear regression model where the null hypothesis that a group of parameter estimates are simultaneously equal to zero can be tested. The test statistic is distributed as chi-square with degrees of freedom equal to the number of restrictions under the null, and a sufficiently large statistic implies a rejection of the null hypothesis. The test was highly significant implying that age2 should be added to the model. Age was expected to have a positive effect initially on the likelihood of a patient receiving an invasive procedure. At some point, elderly patients were anticipated to be less likely to receive invasive procedures implying that the coefficient on age2 would be negative.

Because invasive procedures are relatively more expensive than non-invasive procedures, we might expect that those with higher incomes would be more likely to have invasive procedures. However, this impact will likely be greatly diminished or eliminated for people with adequate insurance coverage. Again, because of the cost, individuals paying the bills should be far less likely to receive an invasive procedure than those with insurance. Since private insurance, Medicare and Medicaid all appear to provide fairly comprehensive coverage, we do

not expect there to be a statistically significant difference between treatment types in these categories. Similarly, elective admissions may be more or less likely to have invasive procedures as their initial treatments. These patients may have been in a better state of health than those that were not admitted electively. Since invasive treatments are usually for more severe cases, patients choosing to enter the hospital may be less likely to have such treatments. But they also may be more motivated to find a cure or more able to afford treatments, thus more likely to have the more expensive invasive treatments. Bedsize was expected to have a slightly positive coefficient as larger hospitals may be better equipped to perform complicated procedures. The number of diagnoses an individual had on his or her discharge record was expected to have a fairly large and positive effect on the probability that an invasive procedure would be carried out. As far back as 1986, Chassin *et al.* (1986) found significant geographic differences in the use of medical and surgical procedures for a wide variety of ailments. Thus, our priors were that we would find geographic differences in hypertension treatments as well. In addition to the independent variables listed above, we attempted to estimate the effect that physicians had on treatment choice. However, after sorting data by attending physician's identifying number, it became apparent that there were a vast number of physicians and very few observations per physician. Owing to the lack of variation in the treatments associated with any given physician, we would not be able to accurately predict the effect that a he or she had on a chosen treatment. We tried testing for interactions among age, race, income, insurance coverage and gender. However, this introduced significant multi-collinearity and also introduced a greater likelihood of empty cells, diminishing the accuracy of the results. Thus, we used the simpler model reported here. Details of these results are available from the lead author upon request

Results

Logistic regression results are contained in Table 4. The Likelihood Ratio test for the model was highly significant ($p<0.0001$). There was a non-linear relationship between the likelihood of a patient receiving surgery and the patient's age. The probability of having an invasive procedure increases up to age 30 ($p<0.0001$), then declines marginally for each subsequent year of age ($p<0.0001$). Females were less likely to receive an invasive procedure ($p<0.0001$).

Most minorities were statistically more likely to have invasive procedures when compared to whites. The increase was strongest for Hispanics ($\beta=0.4444$, $p<0.0001$), then Asians ($\beta=0.04163$, $p=0.0323$), and then Blacks ($\beta=0.1114$, $p=0.0221$). A Wald test shows that these effects are also statistically different from each other ($p=0.0001$).

Every state in the regression, except Utah and Colorado, had a statistically significant influence on treatment choice. The effects associated with most of the states were also quite strong. With the exception of Arizona, patients in these states were less likely to receive an invasive procedure than patients in Florida. The effects were strongest for New Hampshire and Vermont. Using the Wald test again, we determined that these estimates were also significantly different from each other ($p<0.0001$).

When we look at economic variables we see that patients from middle and upper income zip codes had an increased likelihood of receiving invasive procedures. As predicted, "self-pay" ($p=0.0002$) and "otherpay" ($p=0.0061$) patients were less likely to have invasive procedures as initial treatments and these effects were strong ($\beta= -0.3817$ and -0.3823, respectively). Medicare coverage was the only type of insurance positively related with a higher probability of invasive treatments ($p<0.0001$) with another strong effect ($\beta=0.6518$). The estimated coefficients for

elective admission and bed-size were both significantly positive ($p<0.0001$ for both) and the effect of elective was one of the strongest of all qualitative variables.

Discussion and conclusions

When analyzing the initial procedures received by patients with "Hypertension with Complications and Secondary Hypertension" it was clear that treatment alternatives could be classified into two types: (i) invasive and (ii) non-invasive nature. Our results indicate that there are a number of non-clinical factors that play a significant role determining which of these two treatment types an individual is likely to receive. The likelihood of a patient having an invasive procedure performed as his or her primary procedure significantly declined after an early age, which is consistent with previous findings of age being an important predictor of treatment type. The effect of age on treatment choice may be the result of frailty, that is, older patients may face greater risks from invasive treatments. In fact, when this finding is looked at in light of the mean value, the relatively small coefficient of variation and the large range of data indicates that most patients are of an age where the marginal effect of an additional year is negative.

We also confirmed, like the many other studies of other diseases cited above, that race is a significant treatment predictor. All larger minority groups are significantly more likely to have invasive procedures as an initial treatment. It could be that non-whites respond less successfully to non-invasive procedures, so this difference, while interesting to note, does not say anything about the appropriateness of treatments given minority groups.

Compared to patients with private insurance, patients with Medicare had an invasive procedure performed approximately twice as often; "self-pay" and "otherpay" patients were significantly less likely to receive an invasive procedure than individuals with private insurance,

and Medicaid was an insignificant treatment choice determinant. We controlled for age and other socioeconomic characteristics but these results may need further study. More ominously, total charges were significantly higher among patients whose primary procedure was invasive. The average price difference was $9,015.57. Private insurers, Medicaid, self-pay patients and those using a form of payment other than the ones listed above, may have been less willing to pay the costs of using invasive procedures initially. These results could be disheartening if invasive procedures are found to be more effective in treating these types of hypertension, or if they imply that physicians are treating patients differently based on the type of insurance coverage they possess.

Patients living in the middle and upper income zip codes were more likely to have an invasive procedure performed than patients from the lower income zip codes. These results are consistent with previous studies (again, of other diseases) that found higher income increases the likelihood of patients receiving relatively more expensive treatments (Bradley, Given, and Roberts, 2002; Desch *et al.,* 1996; Sturm, Meredith, and Wells, 1996). Although the effect is slight, our results are inconsistent with those found by Roetzheim, and others (2000) who found income to be insignificant in determining treatment choice (since our income proxy is a measure of the median income in the patient's home area, the accuracy of these results depends on how close the patient's actual income is to this median value).

As with many of the studies mentioned above, we also found geographical variations in treatment choice. After conducting a thorough review of the regional treatment variation literature and ruling out alternative explanations, Phelps (1992) suggests that the main reason for these observed differences is an incomplete diffusion of information. Unlike many of the aforementioned studies, we found that geography matters at the state level, not at the regional

level. One potential reason for these findings is that policies differ from state to state (Ahronheim *et al.,* 2001). In our study, only Utah and Colorado were equivalent to Florida - our excluded state. Arizona patients were approximately twice as likely to receive an invasive procedure as those in Florida. Individuals in Florida were anywhere from approximately 1.5 to 9 times more likely to receive invasive treatments than patients seeking treatment in the remaining states. If Florida had an older population relative to the rest of the sample, a couple possible reasons for this finding could be that physicians there are more experienced with invasive procedures, and are therefore more likely to perform them. Or it could be that there is a higher demand for these types of procedures among the older population. However, a Kruskal-Wallis rank test revealed that patient's in Florida were not significantly older than those in the rest of the sample ($p=.2782$). Clearly these results are something that state-level policymakers need to be aware.

The outcomes associated with sex, how the patient was admitted to the hospital and hospital size are difficult to interpret. The results indicate that females were 1.2 times less likely to receive an invasive procedure as their primary treatment than males of the same age. This could be due to differences in severity or effectiveness of non-invasive procedures across gender. The significance of gender was not observed in the previous studies. Patients admitted electively were almost three times as likely to have an invasive procedure performed than those admitted by other means. It could be that individuals who are willing to admit themselves into a hospital play a more active and aggressive role in choosing their health care or it may be that they put it off until the condition is severe enough that more aggressive treatments are required. Our finding on hospital size contradicts Satariano, Swanson, and Moll (1992) who found that patients treated in large hospitals were more likely to receive the relatively less-invasive procedure. They attributed this finding to large hospitals having better access to cutting edge treatments. It could

be that this holds true for our study as well, with the newer forms of invasive therapies. All these findings deserve further attention.

Our results coincide with those in the previous literature, finding that variables not directly related to a patient's clinical status play an important role in treatment choice. Some of the individual outcomes, however, contradict literature findings. Politicians and hospital managers may be especially interested in our results, particularly those concerning race, geography and payer information. It is particularly telling that self-pay patients are less likely to receive the more expensive treatment and that patients covered by Medicare are more likely to receive the more expensive treatments. This holds even after controlling age and co-morbidity, strengthening the implication that Medicare patients are for some reason treated differently. Racial disparity and geographical treatment variations are topics already receiving a great deal of attention in the treatment literature. According to our study, they remain issues.

One way of improving our study would be to consider the role of the physician in treatment choice. We used only patient variables. Given sufficient data to look at variability within the patients treated by a single physician possibly would allow us to determine whether physicians are basing their treatment decisions on these non-clinical variables, or if the patient is the primary decision maker. Future research could tie this information into a simultaneous system of equations in order to determine whether patients who received one treatment type versus another were better off in terms of either length of hospital stay, or their disposition status.

References

Ahronheim, J. C., M. Mulvihill, C. Sieger, P. Park, and B. E. Fries. "State Practice Variations in the Use of Tube Feeding for Nursing Home Residents with Severe Cognitive Impairment." *J Am Geriatr Soc* 49, no. 2 (2001): 148-52.

Bergman, L., G. Dekker, F. E. van Leeuwen, S. J. Huisman, F. S. van Dam, and J. A. van Dongen. "The Effect of Age on Treatment Choice and Survival in Elderly Breast Cancer Patients." *Cancer* 67, no. 9 (1991): 2227-34.

Bradley, C. J., C. W. Given, and C. Roberts. "Race, Socioeconomic Status, and Breast Cancer Treatment and Survival." *J Natl Cancer Inst* 94, no. 7 (2002): 490-6.

Chassin, M. R., R. H. Brook, R. E. Park, J. Keesey, A. Fink, J. Kosecoff, K. Kahn, N. Merrick, and D. H. Solomon. "Variations in the Use of Medical and Surgical Services by the Medicare Population." *N Engl J Med* 314, no. 5 (1986): 285-90.

Desch, C. E., L. Penberthy, C. J. Newschaffer, B. E. Hillner, M. Whittemore, D. McClish, T. J. Smith, and S. M. Retchin. "Factors That Determine the Treatment for Local and Regional Prostate Cancer." *Med Care* 34, no. 2 (1996): 152-62.

Glover, M.J., K.J. Greenlund, C. Ayala, and J.B. Croft. "Racial/Ethnic Disparities in Prevalence, Treatment, and Control of Hypertension--United States, 1999-2002." *Morbidity and Mortality Weekly Report* 54, no. 1 (2005): 7-9.

Greenberg, E. R., C. G. Chute, T. Stukel, J. A. Baron, D. H. Freeman, J. Yates, and R. Korson. "Social and Economic Factors in the Choice of Lung Cancer Treatment. A Population-Based Study in Two Rural States." *N Engl J Med* 318, no. 10 (1988): 612-7.

Gujarati, Damodar N. *Basic Econometrics* 3rd ed. New York: McGraw-Hill, 1995.

HCUP. *Clinical Classification Software (CCS) for ICD-9-CM.* Agency for Healthcare Research and Quality, 2000-2003 [cited 08-04-06. Available from www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp.

HCUP. *Design of the Nationwide Inpatient Sample (Nis), 2003* 2006 [cited 03-03-2006. Available from http://www.hcup-us.ahrq.gov/db/nation/nis/nisrelatedreports.jsp.

Krumholz, H. M., M. J. Radford, Y. Wang, J. Chen, A. Heiat, and T. A. Marciniak. "National Use and Effectiveness of Beta-Blockers for the Treatment of Elderly Patients after Acute Myocardial Infarction: National Cooperative Cardiovascular Project." *Jama* 280, no. 7 (1998): 623-9.

Lewis, C. E. "Variations in the Incidence of Surgery." *N Engl J Med* 281, no. 16 (1969): 880-4.

McPherson, K., P. M. Strong, A. Epstein, and L. Jones. "Regional Variations in the Use of Common Surgical Procedures: Within and between England and Wales, Canada and the United States of America." *Soc Sci Med [A]* 15, no. 3 Pt 1 (1981): 273-88.

McPherson, K., J. E. Wennberg, O. B. Hovind, and P. Clifford. "Small-Area Variations in the Use of Common Surgical Procedures: An International Comparison of New England, England, and Norway." *N Engl J Med* 307, no. 21 (1982): 1310-4.

O'Connor, G. T., H. B. Quinton, N. D. Traven, L. D. Ramunno, T. A. Dodds, T. A. Marciniak, and J. E. Wennberg. "Geographic Variation in the Treatment of Acute Myocardial Infarction: The Cooperative Cardiovascular Project." *Jama* 281, no. 7 (1999): 627-33.

Phelps, C. E. "Diffusion of Information in Medical-Care." *Journal of Economic Perspectives* 6, no. 3 (1992): 23-42.

Phelps, C. E., and S. T. Parente. "Priority Setting in Medical Technology and Medical Practice Assessment." *Med Care* 28, no. 8 (1990): 703-23.

Pilote, L., R. M. Califf, S. Sapp, D. P. Miller, D. B. Mark, W. D. Weaver, J. M. Gore, P. W. Armstrong, E. M. Ohman, and E. J. Topol. "Regional Variation across the United States in the Management of Acute Myocardial Infarction. Gusto-1 Investigators. Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Coronary Arteries." *N Engl J Med* 333, no. 9 (1995): 565-72.

Reynolds, K., D. Gu, P. Muntner, X. Wu, J. Chen, G. Huang, X. Duan, P. K. Whelton, and J. He. "Geographic Variations in the Prevalence, Awareness, Treatment and Control of Hypertension in China." *J Hypertens* 21, no. 7 (2003): 1273-81.

Roetzheim, R. G., E. C. Gonzalez, J. M. Ferrante, N. Pal, D. J. Van Durme, and J. P. Krischer. "Effects of Health Insurance and Race on Breast Carcinoma Treatments and Outcomes." *Cancer* 89, no. 11 (2000): 2202-13.

Samet, J., W. C. Hunt, C. Key, C. G. Humble, and J. S. Goodwin. "Choice of Cancer Therapy Varies with Age of Patient." *Jama* 255, no. 24 (1986): 3385-90.

Satariano, E. R., G. M. Swanson, and P. P. Moll. "Nonclinical Factors Associated with Surgery Received for Treatment of Early-Stage Breast Cancer." *Am J Public Health* 82, no. 2 (1992): 195-8.

Singh, Vibhuti N. *High Blood Pressure* eMedicine, 08-22-05 [cited 08-04-06. Available from http://www.emedicinehealth.com/high_blood_pressure/article_em.htm.

Sturm, R., L. S. Meredith, and K. B. Wells. "Provider Choice and Continuity for the Treatment of Depression." *Med Care* 34, no. 7 (1996): 723-34.

Wennberg, J. E., and A. Gittelsohn. "Health Care Delivery in Maine I: Patterns of Use of Common Surgical Procedures." *J Maine Med Assoc* 66, no. 5 (1975): 123-30, 49.

Wennberg, J., and Gittelsohn. "Small Area Variations in Health Care Delivery." *Science* 182, no. 117 (1973): 1102-8.

Wenneker, M. B., and A. M. Epstein. "Racial Inequalities in the Use of Procedures for Patients with Ischemic Heart Disease in Massachusetts." *Jama* 261, no. 2 (1989): 253-7.

Whittle, J., J. Conigliaro, C. B. Good, and R. P. Lofgren. "Racial Differences in the Use of Invasive Cardiovascular Procedures in the Department of Veterans Affairs Medical System." *N Engl J Med* 329, no. 9 (1993): 621-7.

Table 2.1: Data Descriptions

Variables	Descriptions
Invasive	Binary variable, equals 1 if the initial procedure performed on the patient was invasive, as defined by the ICD-9-CM codes.
Age	Age of patient, in years, at the time of admission.
Age2	The square of the "Age" variable
Female	Binary variable, equals 1 if patient is female.
Black	Binary variable, equals 1 if patient's race is Black
Hispanic	Binary variable, equals 1 if patient's race is Hispanic
Asian	Binary variable, equals 1 if patient's race is Asian or Pacific Islander
Otherace	Binary variable, equals 1 if patient's race is something other than White non-Hispanic, Black, Hispanic, Asian, or Pacific Islander
Middle	Binary variable, equals 1 if the median household income for the patient's zip code was in the $36,000 - $44,999 range.
Uppermid	Binary variable, equals 1 if the median household income for the patient's zip code was in the $45,000 - $59,999 range.
Upper	Binary variable, equals 1 if the median household income for the patient's zip code was $60,000 or more
Selfpay	Binary variable, equals 1 if the patient was the expected primary payer of the bill
Medcare	Binary variable, equals 1 if Medicare was the expected primary payer of the bill
Medcaid	Binary variable, equals 1 if Medicaid was the expected primary payer of the bill
Otherpay	Binary variable, equals 1 if the expected primary payer of the bill was someone other than private insurance, the patient, Medicare, or Medicaid
Elective	Binary variable, equals 1 if the patient was admitted electively
Ndx	Total number of diagnoses coded on the patient's discharge record
Bedsize	Categorical variable based on the number of beds a hospital has. 1 = Small, 2 = Medium, and 3 = Large.
Totchg	The total charges incurred by the patient during his or her stay in the hospital
AZ – WI	Binary variables indicating whether the patient was treated in the respective state

Table 2.2: Summary Statistics

Variable	All Patients (17,437 patients)	Invasive Procedure Received (13,804 patients)	Non-Invasive Procedure Received (3,633 patients)
Invasive	0.79 [51.30] (0.41)	1.00 [0] (0)	0 [.] (0)
Age	60.94 [28.58] (17.41)	59.70 [28.87] (17.24)	65.67 [26.31] (17.28)
Age2	4017.10 [50.64] (2034.18)	3860.89 [51.25] (1978.53)	4610.64 [46.23] (2131.34)
Female	0.52 [95.42] (0.50)	0.51 [98.51] (0.50)	0.58 [84.44] (0.49)
Black	0.41 [120.69] (0.49)	0.41 [118.80] (0.49)	0.38 [128.32] (0.48)
Hispanic	0.12 [275.27] (0.32)	0.12 [266.75] (0.33)	0.09 [315.37] (0.29)
Asian	0.02 [801.94] (0.12)	0.02 [784.00] (0.13)	0.01 [883.17] (0.11)
Otherace	0.03 [601.51] (0.16)	0.03 [606.81] (0.16)	0.03 [582.60] (0.17)
Middle	0.24 [176.62] (0.43)	0.25 [174.63] (0.43)	0.23 [184.66] (0.42)
Uppermid	0.20 [202.01] (0.40)	0.19 [204.17] (0.40)	0.21 [194.29] (0.41)
Upper	0.14 [247.51] (0.35)	0.13 [253.18] (0.34)	0.16 [228.52] (0.37)
Selfpay	0.04 [506.19] (0.19)	0.03 [532.66] (0.18)	0.05 [431.78] (0.22)
Medcare	0.63 [75.82] (0.48)	0.64 [75.13] (0.48)	0.62 [78.50] (0.49)
Medcaid	0.13 [258.10] (0.34)	0.13 [257.68] (0.34)	0.13 [259.77] (0.34)
Otherpay	0.02 [687.78] (0.14)	0.02 [698.84] (0.14)	0.02 [650.09] (0.15)
Elective	0.18 [210.57] (0.39)	0.21 [192.40] (0.41)	0.08 [350.87] (0.26)
Ndx	7.43 [42.10] (3.13)	7.36 [42.96] (3.16)	7.71 [38.75] (2.99)
Bedsize	2.58 [25.02] (0.65)	2.60 [24.39] (0.64)	2.50 [27.23] (0.68)
Totchg	35183.87 [141.72] (49863.69)	37062.28 [142.57] (52839.21)	28046.71 [126.63] (35515.48)
AZ	0.01 [864.97] (0.11)	0.02 [795.04] (0.12)	4.13E-03 [1553.28] (0.06)
CO	0.01 [984.71] (0.10)	0.01 [929.37] (0.11)	0.01 [1344.25] (0.07)
CT	0.02 [757.11] (0.13)	0.02 [768.26] (0.13)	0.02 [718.79] (0.14)
FL	0.14 [247.04] (0.35)	0.16 [232.25] (0.36)	0.08 [335.81] (0.27)
HI	0.01 [1191.36] (0.08)	0.01 [1159.06] (0.09)	0.01 [1344.25] (0.070)
IA	2.87E-03 [1864.83] (0.05)	2.46E-03 [2012.53] (0.05)	4.40E-03 [1503.74] (0.07)
IN	0.03 [525.68] (0.18)	0.04 [514.25] (0.19)	0.03 [576.91] (0.17)
KS	0.01 [950.25] (0.10)	0.01 [950.91] (0.10)	0.01 [947.89] (0.10)
MA	0.02 [733.73] (0.13)	0.02 [759.90] (0.13)	0.02 [654.09] (0.15)
MD	0.04 [469.75] (0.20)	0.04 [472.43] (0.20)	0.05 [459.98] (0.21)
MI	0.02 [697.72] (0.14)	0.02 [720.37] (0.14)	0.02 [627.52] (0.16)
MO	0.04 [461.51] (0.21)	0.05 [444.16] (0.21)	0.03 [550.70] (0.18)
NC	0.05 [421.55] (0.22)	0.05 [420.48] (0.23)	0.05 [425.75] (0.22)
NH	1.38E-03 [2693.67] (0.04)	6.52E-04 [3915.21] (0.03)	4.13E-03 [1553.28] (0.06)
NJ	0.06 [386.14] (0.24)	0.05 [420.78] (0.23)	0.10 [302.50] (0.30)
NY	0.13 [254.26] (0.34)	0.12 [275.90] (0.32)	0.20 [198.93] (0.40)
PA	0.07 [368.51] (0.25)	0.07 [376.21] (0.25)	0.08 [342.79] (0.27)
RI	0.01 [1045.79] (0.09)	0.01 [1100.76] (0.09)	0.01 [893.06] (0.11)
SC	0.02 [654.32] (0.15)	0.02 [632.17] (0.15)	0.02 [765.33] (0.13)
TN	0.06 [381.70] (0.25)	0.07 [374.89] (0.25)	0.06 [411.11] (0.23)
TX	0.16 [232.21] (0.36)	0.16 [225.54] (0.37)	0.13 [262.67] (0.33)
UT	3.56E-03 [1674.09] (0.06)	4.06E-03 [1566.90] (0.06)	1.65E-03 [2459.00] (0.04)
VA	0.04 [480.82] (0.20)	0.04 [478.37] (0.20)	0.04 [490.53] (0.20)
VT	1.1E-03 [2951.10] (0.03)	5E-04 [4439.76] (0.02)	3.6E-03 [1668.95] (0.06)
WI	0.02 [748.24] (0.13)	0.02 [782.21] (0.13)	0.02 [650.09] (0.15)

Note: Values represent Means (Standard Deviations) [Coefficients of Variation*100]

Table 2.3: Cross-Tabulation Results

Variable		Invasive	Non-Invasive
Age*** (<.0001)		[115309799.0]	[36723404.0]
Age2*** (<.0001)		[115309799.0]	[36723404.0]
Ndx*** (<.0001)		[118720978.0]	[33312225.5]
Totchg*** (<.0001)		[123045665.0]	[28482271.0]
Female*** (<.0001)	0	6798	1512
	1	7006	2121
Black*** (<.0001)	0	8079	2260
	1	5725	1373
Hispanic*** (<.0001)	0	12103	3301
	1	1701	332
Asian (.1437)	0	13583	3587
	1	221	46
Otherace (0.4689)	0	13439	3529
	1	365	104
Middle*** (.0117)	0	10395	2809
	1	3409	824
Uppermid** (.0312)	0	11133	2872
	1	2671	761
Upper*** (<.0001)	0	11941	3049
	1	1863	584
Selfpay*** (<.0001)	0	13334	3448
	1	470	185
Medcare** (.0226)	0	4980	1385
	1	8824	2248
Medcaid (.7733)	0	11997	3164
	1	1807	469
Otherpay (.2499)	0	13527	3549
	1	277	84
Elective*** (<.0001)	0	10868	3360
	1	2936	273
Bedsize*** (<.0001)	1	1133	389
	2	3200	1031
	3	9471	2213
AZ*** (<.0001)	0	13589	3618
	1	215	15
CO*** (0.0015)	0	13646	3613
	1	158	20
CT (0.3356)	0	13574	3564
	1	230	69
FL*** (<.0001)	0	11645	3337
	1	2159	296

HI	0	13702	3613
(0.2254)	1	102	20
IA**	0	13770	3617
(0.0516)	1	34	16
IN**	0	13301	3527
(0.0339)	1	503	106
KS	0	13653	3593
(0.9707)	1	151	40
MA**	0	13569	3550
(0.0196)	1	235	83
MD	0	13212	3469
(0.5525)	1	592	164
MI**	0	13543	3543
(0.0251)	1	261	90
MO***	0	13138	3517
(<.0001)	1	666	116
NC	0	13065	3443
(0.7677)	1	739	190
NH***	0	13795	3618
(<.0001)	1	9	15
NJ	0	13066	3275
(<.0001)	1	738	358
NY***	0	12201	2900
(<.0001)	1	1603	733
PA***	0	12893	3348
(0.0082)	1	911	285
RI**	0	13691	3588
(0.0174)	1	113	45
SC***	0	13467	3572
(0.0062)	1	337	61
TN**	0	12887	3430
(0.0210)	1	917	203
TX***	0	11536	3173
(<.0001)	1	2268	460
UT**	0	13748	3627
(0.0302)	1	56	6
VA	0	13226	3488
(0.5980)	1	578	145
VT***#	0	13797	3620
(<.0001)	1	7	13
WI***	0	13582	3549
(0.0040)	1	222	84

Values represent number of observations (P-values) [Rank-Sum values]

*Statistical Significance at P = .10

**Statistical Significance at P = .05

***Statistical Significance at P = .01

25% of the cells have expected counts < 5. Chi-Square may not be a valid test.

Table 2.4: Logistic Regression Results

	Dependent Variable		
	Invasive		
Independent Variable	**Coefficients**	**Chi-Square**	**P-Values**
Intercept***	1.258	30.5862	<.0001
Age***	0.0294	20.0205	<.0001
Age2***	-0.00048	75.0171	<.0001
Female***	-0.1851	21.1674	<.0001
Black**	0.1114	5.2379	0.0221
Hispanic***	0.4444	34.1472	<.0001
Asian**	0.4163	4.5809	0.0323
Otherace	0.1152	0.8742	0.3498
Middle**	0.1231	5.3234	0.021
Uppermid	0.0892	2.4991	0.1139
Upper**	0.1276	3.8573	0.0495
Selfpay***	-0.3817	13.8247	0.0002
Medcare***	0.6518	119.2473	<.0001
Medcaid	0.00581	0.0062	0.9371
Otherpay***	-0.3823	7.5262	0.0061
Elective***	1.0831	244.677	<.0001
Ndx	-0.0103	2.1799	0.1398
Bedsize***	0.1516	25.5385	<.0001
AZ**	0.6688	5.7922	0.0161
CO	-0.0494	0.0384	0.8447
CT***	-0.4868	9.4081	0.0022
HI**	-0.5754	3.7874	0.0516
IA***	-1.2417	14.5558	0.0001
IN***	-0.3918	9.3774	0.0022
KS***	-0.637	10.6028	0.0011
MA***	-0.7409	24.8761	<.0001
MD***	-0.5383	22.3171	<.0001
MI***	-0.8126	32.9655	<.0001
MO*	-0.2299	3.5203	0.0606
NC***	-0.6882	42.1893	<.0001
NH***	-2.2178	24.7544	<.0001
NJ***	-1.2447	174.9342	<.0001
NY***	-1.1556	208.3754	<.0001
PA***	-0.7155	55.8816	<.0001
RI***	-0.9158	22.2623	<.0001
SC*	-0.2786	3.1861	0.0743
TN***	-0.5807	31.8056	<.0001
TX***	-0.5342	40.8397	<.0001
UT	0.00912	0.0004	0.9836
VA***	-0.6316	29.2896	<.0001
VT***	-2.1626	19.6401	<.0001
WI***	-0.8665	34.9985	<.0001

*Statistical Significance at p = .10, ** at p = .05, & *** at p = .01

CHAPTER THREE

PATIENTS' PERCEPTIONS AND TREATMENT EFFECTIVENESS

Abstract

Though there is an extensive literature regarding the relationship between patients' expectations and treatment outcomes, studies have failed to address the determinants of these expectations. We argue that factors such as treatment history provide a reference point that may influence patients' expectations of how effective further treatment might be. Therefore, decisions about whether to proceed with additional treatment, as well as perceptions of how effective that treatment is, may be influenced by this frame of reference. We expect that patients with unsuccessful prior treatments have a frame of reference leaving them less likely to expect improvement from subsequent treatments. Prospect theory is used to develop a theoretical foundation for these frame of reference effects on expectations and subsequent treatment. Using data on patients diagnosed with idiopathic intracranial hypertension we test for a frame of reference effect. The results support the proposition that prior treatment failure or success impacts a patient's reference point, which in turn influences her perceptions about the effectiveness of subsequent treatments.

Keywords: Prospect Theory, Expectations, Treatment outcomes, Treatment history, logistic regression, Monotone rank estimator.

Introduction

For some illnesses or health disorders treatments are palliative rather than curative. When conditions are without clear objective symptoms one commonly used measure of treatment effectiveness is self-reported changes in disease status. Prior experience, particularly with earlier treatment for the condition, provides a reference point that may affect a patient's perception of how much further treatment will improve her health status. More specifically, by affecting patients' expectations about how effective a treatment might be, this reference point influences whether a patient pursues an additional treatment, as well as perceptions of how effective that treatment is. Although there is an extensive literature on how patient expectations influence treatment outcome and recovery speed, to our knowledge no existing research tests how expectations depend on prior treatment and other personal characteristics. The purpose of this paper is to explore how reference points affect the perceived effectiveness of medical treatments.

The literature regarding the relationship between expectations and health is extensive, but many of these studies focus only on the fact that a connection exists, and not on the causal relationship (Carver et al., 1994; Frey et al., 1985; Koller et al., 2000).[1] For example, Miceli and Castelfranchi (2002) speculate on the psychological effects of combining forecasts of future events with hopes and fears, both before and after the event had occurred.

Another line of research found that positive expectations speeded recovery (Scheier, 1989 and Scheier and Carver, 1987 for coronary bypass surgery; Frey et al., 1985 for recovery from accidents; and Kalauokalani et al., 2001, for low back pain). Others have found that positive expectations improve patients' perceptions of subjective health (Carver et al., 1994 for breast cancer patients; Llewellyn-Thomas, Thiel, and McGreal, 1992 for the general assessment of one's own health; and Koller et al., 2000 for the quality of life of cancer patients).

[1] For summaries see: Ditto and Hilton (1990), Jones (1982), and Jones (1990).

But none of these studies address what determines expectations of treatment outcomes. We argue that prior treatment provides a frame of reference which, along with other personal characteristics, affects patients' baseline expectations of treatment success, which in turn influences their perceived effectiveness of that treatment. We expect that patients with unsuccessful prior treatments have a frame of reference leaving them less likely to expect improvement from subsequent treatments. We use Prospect Theory (Kahneman and Tversky, 1979) as a basis for a theoretical foundation for these frame of reference effects on expectations and subsequent treatment. We test these ideas using the Monotone Rank Estimator (MRE) (Cavanagh and Sherman, 1998) with data on patients diagnosed with idiopathic intracranial hypertension (IIH).[2] The results support the proposition that prior treatment failure or success impacts a patient's reference point, which in turn influences her perceptions about the effectiveness of subsequent treatments.

The remainder of this paper is organized into four sections. In the next section we describe our model of treatment choice. Successive sections discuss our data and our empirical model and estimation results. We finish the paper with conclusions and implications for future research.

Theory

Prospect theory (PT) was introduced as an alternative to expected utility theory (Von Neumann and Morgenstern, 2004) for modeling decisions under risk when those decisions are dependent on a frame of reference.[3] Unlike expected utility theory where values are placed on final states,

[2] See Appendix A for a detailed explanation of this disorder.
[3] A revised version titled "cumulative prospect theory" (Tversky and Kahneman, 1992; Wakker and Tversky, 1993), applies to uncertain and risky prospects with multiple outcomes. Under the extended theory decision weights are applied to cumulative, as opposed to individual, probabilities. However, the extension gives the same results as the

PT assumes that individuals assign values to gains and losses relative to a reference point. The frame of reference can in principle influence the valuation of possible outcomes, the subjective probabilities of treatment effectiveness, and risk preferences. In our application, the patient's reference point affects her expectations of treatment success, thus determining the path of treatment, and, for subjective outcomes, the perceived success of treatments after the treatments have been pursued.

The value of a given prospect is measured by:

$$V(x, p; y, q) = \pi(p)v(x) + \pi(q)v(y), \quad (1)$$

where x and y are potential outcomes that occur with probabilities p and q, respectively. The decision weight, $\pi(\cdot)$, measures not only the impact of the perceived (as opposed to actual) probabilities on the overall valuation of the prospect, but also the influence of factors such as ambiguity. Kahneman and Tversky (1979) argue that in most cases $\pi(p) < p$, and $\pi(p) + \pi(1-p) < 1$. However, they also argue that small probabilities tend to be overweighted so that $\pi(p) > p$ if p is small.

The value function $v(\cdot)$ measures the value of gains and losses relative to the reference point. This function is believed to be concave for gains and convex for losses, giving it an S-shape, as illustrated in Figure 1 (Kahneman and Tversky, 1979). Also note that the function passes through the reference point, and is steeper for losses than gains (risk aversion). The overweighting of low probability events and the underweighting of high probability events is what can cause some people to be risk seeking for potential losses and risk averse for potential gains.

---

original theory for all two-outcome and mixed three-outcome prospects. This paper focuses on three-outcome prospects, hence the earlier version of Prospect Theory (Kahneman and Tversky, 1979) is sufficient for our purposes.

Though PT has been applied to topics in the field of health, its uses have been relatively limited. One of the first and primary applications focuses on the effects of information framing on medical decisions (McNeil, Pauker, and Tversky, 1988). Meyerowitz and Chaiken (1987) find that the use of negatively framed information leads to increased breast self-examination. Rothman et al. (1993) further corroborate the importance of framing by demonstrating that negative framing may be more effective in encouraging behaviors that are seen as risky (for example being tested for a sexually transmitted disease), while positive framing may be more effective in encouraging preventive behavior (for example, practicing safe sex). It has also been shown that negative (positive) framing may be more persuasive when the perceived efficacy of a solution is low (high) (Block and Keller, 1995).

Lenert, Treadwell, and Schwartz (1999) provide empirical support for the "S" shaped utility function for health. Treadwell and Lenert (1999) propose that the individual's current health state is the reference level. In more recent work, Rasiel, Weinfurt, and Schulman (2005) use PT to rationalize risk-seeking behavior among terminally ill patients. They suggest that patients' reference points differ due to factors such as pre- and post-diagnosis life expectancies which therefore affect the chosen treatment paths.

Unlike traditional PT models, which have the reference point affecting the valuation of outcomes, our model has the patient's reference point influencing the subjective expectations of treatment success, thus determining the path of treatment, and impacting the perceived success of treatments after they have been pursued.

A Prospect Theory Model of Treatment Choice

In our model patients choose a treatment path if they expect it to have a positive impact on their current health status. We assume that an individual's expectations of post-treatment disease status depend on her subjective probabilities of treatment effectiveness, which in turn are influenced by her reference point. The reference point for our purposes is the patient's current status and her success with earlier treatments. We assume there are three possible outcomes of treatment: the patient improves (b = "better") , the patient remains the same (0 = "no change"), and the patient worsens (w = "worse").

These assumptions give us the following equation:

$$V(b, p; w, q; 0, r) = \pi(p)v(b) + \pi(q)v(w) + \pi(r)v(0). \qquad (2)$$

The value *v(b)* is the value a patient places on feeling better relative to her current status, which occurs with objective probability *p.* This is a gain in well-being, so *v(b)* is positive. On the other hand, *v(w)* is the value placed on feeling worse relative to her current status, and occurs with probability *q.* This is a loss in well-being and therefore *v(w)* is negative. A third possibility is that the patient observes no change from her current situation with a resulting value of *v(0)*, and probability *r=1-p-q*.

One might initially expect *v(0)* to equal 0, but values are affected by the reference value, and as a result *v(0)* could be negative if the current situation is relatively poor compared to earlier states, although $v(0)$ would still exceed $v(w)$. We hypothesize that $0>v(0)>v(w)$ because people seek medical treatment to improve their condition contingent upon the outcomes of previous treatments. We further maintain that losses are feared at least as much as gains are valued, so the absolute value of $v(w)$ will equal or exceed that of $v(b)$ given a one unit change in disease status.

The decision weights, $\pi(\cdot)$, are the subjective probabilities of each outcome. Thus (2) represents the individual's subjective valuation of the treatment outcome. As such, a negative (positive) overall value indicates that the individual perceives a negative (positive) valuation of pursuing further treatment. A zero value indicates a perceived neutral valuation of receiving subsequent treatment. To a large extent the final value will depend on the relative magnitudes of the decision weights. For example, we know that if a patient anticipates an equal one unit change in disease status one way or the other, i.e. $b = -w$, if $\pi(p) \leq \pi(q)$ and if $v(0) = 0$ or $\pi(r)=0$ or both, the overall prospect valuation from (2) would be negative due to the absolute valuation of a loss being at least equal to that of a gain.

Assuming medical treatment is voluntary, all individuals agreeing to a treatment should have a positive subjective valuation at their reference point of the outcome; that is, $V(b,p;w,q;0,r) = \pi(p)v(b)+\pi(q)v(w)+\pi(r)v(0) > 0$. Assuming that $v(0)=0$ or $\pi(r)=0$ or both, this means $\pi(p)v(b) > -\pi(q)v(w)$, or the subjective weighted gain from improving b units must exceed the subjective weighted loss from regressing w units.[4] Because $v(w)$ is negative, the right hand side of the inequality is positive. The inequality can then be rearranged to

$$\pi(p)/\pi(q) > -v(w)/v(b). \qquad (3)$$

Equation 3 implies that a treatment is pursued only if the subjectively weighted probability of gain relative to that of a loss exceeds the ratio of the (absolute) utility value of a loss divided by the utility value of a gain. In terms of our specific application this has an interesting interpretation. For any given values of $v(w)$ and $v(b)$, treatment is more likely to be pursued if

[4] Even though the model is conditioned on the assumption that v(0) = 0, it is a trivial extension to extend the results to the non-restricted case. The math is available upon request from the authors. Also, as we mentioned above, v(0) is most likely less than zero, otherwise treatment would not be sought. Therefore, this assumption essentially allows us to group v(0) and v(w) and analyze the probability of feeling better relative to not.

there is only a relatively small subjective probability of the treatment leading to a worse outcome.

In Appendix A we discuss how the treatment path for IIH follows a specific sequence. Acetazolamide is the drug generally viewed as the most effective form of treatment for IIH, and is usually the treatment of first resort. Other medicinal options are generally reserved for patients who cannot tolerate acetazolamide. If medication is ineffective, subsequent treatment normally involves an invasive procedure such as neurosurgical shunts. Consider two types of individuals, keeping in mind that treatment paths are being evaluated at a point in time after the previous treatment results have been observed. Also assume for simplicity that all individuals place the same value on a one unit deviation from the reference point, regardless of what that point is. Patient A is initially prescribed a medication of some sort; however, due either to its ineffectiveness or the individual's inability to tolerate it, the patient is then given an invasive procedure. Patient B on the other hand, for some reason moves directly to an invasive procedure. We expect that because they have different frames of reference due to different treatment paths, these two individuals will have different perceived valuations for the invasive procedure. Given that A has already experienced a failed treatment we anticipate that person B would have a higher perceived probability ratio of success to failure. This implies the following:

$$[\pi(p)/\pi(q)]_B > [\pi(p)/\pi(q)]_A. \qquad (4)$$

One testable hypothesis implied by (4), but for which we unfortunately do not have the right data, is that prior failed treatment makes it less likely that individuals will pursue subsequent treatment. However, by extension if, as the existing literature contends expectations directly influence perceived treatment outcomes, a testable hypothesis implied by equation (4) is

that patient A will be less likely to report an improvement in disease status from the invasive treatment, having already suffered a failure.

Hypothesis 1: All else equal, a patient who has received more (fewer) failed treatments will be less (more) likely to report an improvement from the latest treatment.

In our application to IIH, hypothesis 1 suggests that among patients who received a medicinal prescription other than acetazolamide, those who had an initial unsuccessful experience with acetazolamide should be less likely to report an improvement in disease status than those who only had the alternative prescription.[5] Furthermore, among patients who eventually received an invasive procedure, those who received only the invasive procedure will be most likely to report an improvement, followed by patients who first had an alternative prescription, then individuals who were treated with acetazolamide before receiving the procedure, and finally patients who had both acetazolamide and an alternative prescription before receiving an invasive procedure. The reason we expect patients who had acetazolamide prior to an invasive procedure to have a lower probability ratio than those who had an alternative prescription before a procedure is because acetazolamide is believed to be the most effective form of therapy for IIH. We therefore expect patients who experienced a failure of this drug to be even less likely to anticipate an improvement from subsequent treatment. To summarize: patients with fewer failed treatments will be more likely to report an improvement in the latest treatment than those having experienced more failed treatments.

[5] In our application to IIH our ordering assumes that the acetazolamide was ineffective from its initial use. If acetazolamide was at first effective and improving a patient's condition, but over time lost its effectiveness, the ordering of anticipated probabilities may be opposite what we discuss.

Data

Data for this study come from the Intracranial Hypertension Registry.[6] The registry gathers information from individuals diagnosed with intracranial hypertension and their physicians. Patients are admitted to the registry on a voluntary basis, but cooperation from at least one of the patient's physicians is required.

Because participation in the registry is voluntary and either self or physician initiated, the patients in the registry may not be representative of the entire population of IIH sufferers. At the time of this study the registry contained information from 732 IIH patients. This study focuses solely on patients in the registry who reported a disease status relative to their pre-diagnosis condition. Individuals who did so rated their relative health status on a scale of 0 to 10, 5 being "no change". Our primary variable of interest, *better*, is a binary variable created from this scale. Patients who had a rating between 6 and 10 were given a value of 1, and based on the assumption that patients would most likely place a negative value on observing no change in their disease status, patients who claimed a status of "no change" or "worse" are grouped into a "not-better" category and coded as 0. One hundred fifty-one observations remained after deleting missing values, the majority resulting from patients who did not report a post-treatment disease status. Descriptions of the variables used in this study are contained in Table 1.

Patient socioeconomic variables include real income earned from the last year worked *(Earnings)* and a collection of binary variables: whether the patient has health insurance *(Health_Ins),* whether the patient is working *(Working),* whether or not the patient is in a medical profession *(MedDv),* whether the patient has vision problems, *(Vision),* is obese *(Obese),* or

[6] The registry is co-sponsored by the Intracranial Hypertension Research Foundation of Vancouver, Washington and the Casey Eye Institute at the Oregon Health and Science University (OHSU). http://www.ihrfoundation.org/news/registry.asp

suffers from headaches *(Headache).* *Male* and *White* are dummy variables indicating the gender and race of the patient.

Each patient was placed into one of 8 treatment categories. *AcetaRx* indicates that the patient had been prescribed acetazolamide as well as an alternative medication. *AcetaInv* and *RxInv* identify patients that received an invasive procedure in addition to acetazolamide or an alternative medication, respectively. *Invasive, Rx,* and *Aceta* are all binary variables which equal one if the patient received only the given treatment. *AcetaRxInv* indicates whether the patient received all three forms of treatment. Patients who did not receive any of the above treatments serve as the base case. A dummy variable was also created to identify patients who attempted weight reduction (*Diet*), because diets are often recommended by physicians due to the apparent link between obesity and IIH.

Summary statistics for the sample and the registry can be found in Table 2. As shown by t-tests indicated in Table 2, the sample does not appear to be representative of all patients in the registry. Most mean values differ at p-values of 0.10 or less. However, our sample does appear to be fairly representative of the population of all IIH patients, the statistics for which are also presented in Table 2. Sixty-seven percent of the patients in our sample experienced vision problems, 85% suffered from headaches at some point, and 49% were obese at the time of diagnosis. Roughly 93% of our sample is female and 92% is white.

Forty-seven percent of our sample had attempted weight reduction (*Diet*), 9% had been prescribed only acetazolamide (*Aceta*), 5% had only been on an alternative form of medication (*Rx*), and 25% only received an invasive procedure (*Invasive*). Five percent of our sample had been on both acetazolamide and an alternative prescription (*AcetaRx*), 21% had taken both of these and received an invasive procedure (*AcetaRxInv*), 14% had been on acetazolamide and

received an invasive procedure (*AcetaInv*), while 17% had taken an alternative form of medication and had an invasive procedure (*RxInv*).

Empirical Methods

Our model implies that an individual pursues a subsequent treatment only if her ratio of the subjectively weighted probability of a gain relative to that of a loss exceeds the ratio of the (absolute) utility value of a loss divided by the utility value of a gain, and that, in turn, influences a patient's probability of assessing a taken treatment as effective, leading to hypothesis 1 that patients with fewer failed treatments are more likely to report an improvement from their latest treatment. The empirical results are conditioned on a point in time after treatment paths have been observed, so the treatment paths can be represented in the model by dummy variables. Our dependent variable is the patient's self-assessed post-treatment disease status. Because this binary variable is created by imposing a chosen cutoff point on the 0-10 scale variable measuring the patient's perceived health status,[7] there may be some miscoding, and thus misclassification, of the dependent variable. Therefore, the variable measuring the patient's true subjective outcome from the treatment is latent. The relationship between the patient's subjective outcome and the observed response is

$$\Pr\left[\left(\frac{\pi(p)}{\pi(q)} > \frac{v(w)}{v(b)} \,\middle|\, \boldsymbol{z}\right)\right] \Rightarrow \Pr(T=1) \rightarrow R\begin{bmatrix}1\\0\end{bmatrix} \Rightarrow R^*\begin{bmatrix}1\\0\end{bmatrix}, \tag{7}$$

where $\boldsymbol{z}$ is a vector of patient characteristics, including treatment history, T=1 implies that a given treatment was chosen, and R is the measured result. R^* represents the latent dependent variable denoting true subjective gains.

[7] See the discussion of the dependent variable in the Data section.

Ordinarily a binary logistic or probit regression would be performed in this case to analyze the probability of a patient feeling *better* relative to *not-better*. However, as mentioned above, one potential problem with this model is that the dependent variable is subject to misclassification error. Failure to control for this when estimating a discrete-response model via traditional techniques such as Logit or Probit, can result in inconsistent estimates (Hausman, Abrevaya, and Scott-Morton, 1998). Abrevaya and Hausman (1999) recommend using the semiparametric MRE as an alternative to parametric estimation. Unlike the parametric approach, semiparametric estimation does not require that the mismeasurement be modeled correctly in order to obtain consistent estimates. Therefore, the MRE is used to adjust for the potential misclassification bias.[8]

For each model, we estimate the probability that a status of *better* occurs as a function of a linear index in the following patient characteristics and treatment variables: *Male, White, Earnings, Health_Ins, Working, MedDv, Vision, Obese, Headache, Diet, Rx, Aceta, Invasive, AcetaRx, AcetaRxInv, RxInv, and AcetaInv.*

To test hypothesis 1, the effects of the relevant treatment paths on the dependent variable are tested against each other. A Wilcoxon signed rank sum test, the nonparametric alternative to the paired t-test, is performed on the $\boldsymbol{X\hat{\beta}}$ vectors from two regressions, one where the two treatment variables being tested enter the regression separately and one where they are combined to form one variable. This test consists of ranking the absolute differences between each pair of $X_i\hat{\beta}$s and calculating the Wilcoxon signed rank statistic to test the whether the median difference is zero. If the restriction is valid, the sum of the ranks for the positive differences should approximately equal the sum of the ranks of the negative differences. A rejection of the

[8] See Appendix B for a detailed discussion of the MRE.

null hypothesis that the median $X_i\hat{\beta}$s are equal, i.e. $X_i\hat{\beta}^{MRE}_{Unrestricted} = X_i\hat{\beta}^{MRE}_{Restricted}$, implies that the effects of the two treatment variables being tested are significantly different, and thus can be compared to one another.

Results

Results from the MRE model can be viewed in Table 3. Health insurance (*Health_Ins*) has a positive influence on disease status, and its estimate is the largest among all explanatory variables. Patients in the medical profession (*MedDv*) and those who are obese are also significantly more likely to report an improvement in disease status. The remaining significant patient characteristic variables have negative effects on the probability of *better* being reported by the patient, they are: *Male, White, Earnings, Working, Vision,* and *Diet.*

In terms of the seven primary treatment variables, we are most interested in comparing the magnitudes of the effects that the previous treatments have on the probability that a patient claims a status of *better* relative to *not better*.[9] Patients who received only acetazolamide (*Aceta*) show a highly significant increase in the probability of reported improvement. This group has the largest coefficient estimate of all the treatment variables. Individuals who had taken acetazolamide as well as a different medication (*AcetaRx*) have the second largest estimate, followed by those who only had an alternative medication (*Rx*), then patients who received acetazolamide as well as an invasive procedure (*AcetaInv*), those who only had an invasive procedure (*Invasive*) are next, then patients who received a medication other than acetazolamide in addition to an invasive procedure (*RxInv*), and finally patients who received all three forms of

[9] This is possible because all of these variables are binary and share the same omitted category.

treatment (*AcetaRxInv*). All of these estimates are positive. The expected and actual orderings of the treatment variables are displayed in Table 4.

Discussion

Our primary hypothesis is that patients with unsuccessful prior treatments have a frame of reference leaving them less likely to expect improvement from subsequent treatments, and therefore less likely to report an improvement in disease status. Our results indicate two failures out of six tests of this hypothesis: using acetazolamide prior to another treatment appears to increase a patient's reported improvement from the subsequent treatment. These results are presented in Table 4. Patients who had acetazolamide prior to receiving a different prescription are more likely to claim a status of *better* than those who only had the alternative medication. The same is true of patients who had acetazolamide and an invasive procedure versus those who bypassed the medication and went straight to an invasive procedure. This might indicate that for many of the patients acetazolamide was initially effective, giving them a frame of reference that the disease symptoms are treatable, but that they were unable to tolerate acetazolamide for an extended period of time due to its side effects (Wall, n.d.; IHRF, 2007c). The ordering suggested in equation 4 appears to hold for the remainder of the patients. That is, patients who only had an invasive procedure are most likely to report an improvement, followed by patients who first had an alternative prescription, and finally, patients who had both acetazolamide and an alternative prescription before receiving an invasive procedure.

Except for the case of acetazolamide we have support for our primary hypothesis, and that the deviation with the case of acetazolamide could be due to its outstanding effectiveness at reducing CSF within the skull (Gücer and Viernstein, 1978; Lubow and Kuhr, 1976; Rubin et al.,

1966; Tomsak, Niffenegger, and Remler, 1988). Due to acetazolamide's effectiveness, patients who were first on that drug may be predisposed to other treatments being effective, thereby increasing their expectations of success for subsequent treatments, and in turn increasing the likelihood of perceiving the given treatment to be effective. Further support for this conjecture comes from the fact that the treatment group consisting of patients who were only prescribed acetazolamide is by far the most likely to report an improvement in disease status. The coefficient estimate for this group is over twice as large as the estimate for patients who in addition to having had acetazolamide were prescribed an alternative form of medication.

Two additional insights about the treatment of IIH come out of our analysis. The first is that attempted weight reduction is the only form of treatment to decrease the probability that a patient would end up in the *better* category. When one considers how difficult dieting is for most people and the low success rate that exists among dieters, this result is not surprising.[10] Dieters are frustrated with the unsuccessful results leaving them with lower expectations for the treatment's effectiveness.

Perhaps our finding that has the most important policy implications is that patients with health insurance are significantly more likely to see an improvement than those without it. This may be due to the fact that the nature of the disease leads physicians to treat symptoms rather than the root cause, which is expensive, and those with health insurance would be more likely to be able to afford an extensive set of treatments over an extended period of time. [11]

Among the other variables in the model, higher earnings decreases the probability of a *better* status; however, the coefficient estimate is close to zero; individuals with a history of vision problems and those who were working prior to IIH interfering with their daily life are

[10] The long term success rate among all dieters is only 31%. For females the success rate is only 27%. (Kruger, Blanck, and Gillespie, 2006).
[11] See Tanne, Rosenman, and Friesner (2008) for a detailed discussion on the economic costs of IIH.

significantly less likely to report an improvement in post-treatment disease status, as are males and whites; while patients in the medical profession and those who are obese are more likely to report an improvement. These results are difficult to explain; however, they may deserve further attention.

Conclusions

We provide a conceptual basis for how prior treatment failure or success along with other factors can influence a patient's reference point that helps determine expectations and decisions on future treatment, and empirically demonstrate support that a patient's prior treatment failure or success may impact her perceptions about the effectiveness of subsequent treatments, and identified other factors that might influence patients' reference points and perceptions. The finding that prior treatment results affect the perceived success of subsequent treatments is important for physicians to know when prescribing additional treatment. Physicians play a large role in providing information, and this knowledge would be very useful when prepping patients for whether or not the subsequent treatment will be effective. The health insurance finding suggests that it may be especially important for patients diagnosed with idiopathic disorders where treatment is focused on symptoms to have good health insurance coverage if they are going to experience at least a perceived improvement in health status.

One of the primary shortcomings of our analysis is that our data come from a voluntary, self-reported registry. While the data are likely to be quite accurate, they may not be representative of the population of IIH sufferers as a whole. As such, this potential registration bias may limit some of our findings, particularly those associated with the race and gender of the patients. Additionally, we are not able to explicitly control for the costs of IIH, which may

influence the decision making process, nor does the study account for all other comorbidities that might occur because of IIH. Finally, our data limitations require an empirical analysis that relies on indirect evidence. Our empirical approach requires we assume that prior expectations of how successful a treatment will be are matched by *ex post* perceptions of treatment effectiveness. A more direct test would utilize the implications from equation (4) and compare directly how prior unsuccessful (or successful, for that matter) treatment influences patients' perceptions of expected success of subsequent treatment, and therefore, their propensity to have the treatment.

References

Abrevaya, Jason, and Jerry A. Hausman. "Semiparametric Estimation with Mismeasured Dependent Variables: An Application to Duration Models for Unemployment Spells." *Annales D'Economie et de Statistique* 55-56 (1999): 243-75.

Ball, A. K., and C. E. Clarke. "Idiopathic Intracranial Hypertension." *Lancet Neurol* 5, no. 5 (2006): 433-42.

Beecher, H. K. "The Powerful Placebo." *J Am Med Assoc* 159, no. 17 (1955): 1602-6.

Binder, D. K., J. C. Horton, M. T. Lawton, and M. W. McDermott. "Idiopathic Intracranial Hypertension." *Neurosurgery* 54, no. 3 (2004): 538-51; discussion 51-2.

Bleichrodt, H., and J. L. Pinto. "A Parameter-Free Elicitation of the Probability Weighting Function in Medical Decision Analysis." *Management Science* 46, no. 11 (2000): 1485-96.

Block, Lauren G., and Punam Anand Keller. "When to Accentuate the Negative: The Effects of Perceived Efficacy and Message Framing on Intentions to Perform a Health-Related Behavior." *Journal of Marketing Research* 32, no. 2 (1995): 192-203.

Bynke, G., G. Zemack, H. Bynke, and B. Romner. "Ventriculoperitoneal Shunting for Idiopathic Intracranial Hypertension." *Neurology* 63, no. 7 (2004): 1314-6.

Camerer, Colin F. "Prospect Theory in the Wild: Evidence from the Field." In *Choices, Values, and Frames*, edited by Daniel Kahneman and Amos Tversky, xx, 840 p. Cambridge, UK: Cambridge University Press, 2000.

Carver, C. S., C. Pozo-Kaderman, S. D. Harris, V. Noriega, M. F. Scheier, D. S. Robinson, A. S. Ketcham, F. L. Moffat, Jr., and K. C. Clark. "Optimism Versus Pessimism Predicts the

Quality of Women's Adjustment to Early Stage Breast Cancer." *Cancer* 73, no. 4 (1994): 1213-20.

Cavanagh, C., and R. P. Sherman. "Rank Estimators for Monotonic Index Models." *Journal of Econometrics* 84, no. 2 (1998): 351-81.

Curry, W. T., Jr., W. E. Butler, and F. G. Barker, 2nd. "Rapidly Rising Incidence of Cerebrospinal Fluid Shunting Procedures for Idiopathic Intracranial Hypertension in the United States, 1988-2002." *Neurosurgery* 57, no. 1 (2005): 97-108; discussion 97-08.

Ditto, P. H., and J. L. Hilton. "Expectancy Processes in the Health-Care Interaction Sequence." *Journal of Social Issues* 46, no. 2 (1990): 97-124.

Frey, D., O. Rogner, M. Schuler, C. Korte, and D. Havemann. "Psychological Determinants in the Convalescence of Accident Patients." *Basic and Applied Social Psychology* 6, no. 4 (1985): 317-28.

Friedman, D. I., and D. M. Jacobson. "Idiopathic Intracranial Hypertension." *J Neuroophthalmol* 24, no. 2 (2004): 138-45.

Friedman, Deborah I. "A Practical Approach to Intracranial Hypertension." *Headache Currents* 2, no. 1 (2005): 1 - 10.

Gans, Mark. "Idiopathic Intracranial Hypertension." In *eMedicine*, 2005.

Giovannini, Joseph, and Georgia Chrousos. "Papilledema." In *eMedicine*, 2005.

Goodwin, James. "Pseudotumor Cerebri." In *eMedicine*, 2006.

Gucer, G., and L. Viernstein. "Long-Term Intracranial Pressure Recording in the Management of Pseudotumor Cerebri." *J Neurosurg* 49, no. 2 (1978): 256-63.

Han, A. K. "Nonparametric Analysis of a Generalized Regression-Model - the Maximum Rank Correlation Estimator." *Journal of Econometrics* 35, no. 2-3 (1987): 303-16.

Hausman, J. A., J. Abrevaya, and F. M. Scott-Morton. "Misclassification of the Dependent Variable in a Discrete-Response Setting." *Journal of Econometrics* 87, no. 2 (1998): 239-69.

IHRF. *Causes of Secondary Ih* Intracranial Hypertension Research Foundation, 2007 [cited. Available from http://www.ihrfoundation.org/intracranial/hypertension/info/C64.

———. *Idiopathic Ih and Weight* Intracranial Hypertension Research Foundation, 2007 [cited. Available from http://www.ihrfoundation.org/intracranial/hypertension/info/C64.

———. *Medication and Surgery* Intracranial Hypertension Research Foundation, 2007 [cited. Available from http://www.ihrfoundation.org/intracranial/hypertension/info/C64.

———. *What's the Difference between Pseudotumor Cerebri, Idiopathic Intracranial Hypertension and Secondary Intracranial Hypertension?* Intracranial Hypertension Research Foundation, 2007 [cited. Available from http://www.ihrfoundation.org/intracranial/hypertension/info/C64.

Johnson, L. N., G. B. Krohel, R. W. Madsen, and G. A. March, Jr. "The Role of Weight Loss and Acetazolamide in the Treatment of Idiopathic Intracranial Hypertension (Pseudotumor Cerebri)." *Ophthalmology* 105, no. 12 (1998): 2313-7.

Jones, R. A. "Expectations and Delay in Seeking Medical-Care." *Journal of Social Issues* 46, no. 2 (1990): 81-95.

———. "Expectations and Illness." In *Interpersonal Issues in Health Care*, edited by Howard S. Friedman and M. Robin DiMatteo, xvii, 301 p. New York: Academic Press, 1982.

Kahneman, D., and A. Tversky. "Prospect Theory - Analysis of Decision under Risk." *Econometrica* 47, no. 2 (1979): 263-91.

Kalauokalani, D., D. C. Cherkin, K. J. Sherman, T. D. Koepsell, and R. A. Deyo. "Lessons from a Trial of Acupuncture and Massage for Low Back Pain: Patient Expectations and Treatment Effects." *Spine* 26, no. 13 (2001): 1418-24.

Koller, M., W. Lorenz, K. Wagner, A. Keil, D. Trott, R. Engenhart-Cabillic, and C. Nies. "Expectations and Quality of Life of Cancer Patients Undergoing Radiotherapy." *J R Soc Med* 93, no. 12 (2000): 621-8.

Kruger, J., H. M. Blanck, and C. Gillespie. "Dietary and Physical Activity Behaviors among Adults Successful at Weight Loss Maintenance." *Int J Behav Nutr Phys Act* 3 (2006): 17.

Kupersmith, M. J., L. Gamell, R. Turbin, V. Peck, P. Spiegel, and M. Wall. "Effects of Weight Loss on the Course of Idiopathic Intracranial Hypertension in Women." *Neurology* 50, no. 4 (1998): 1094-8.

Lenert, L. A., J. R. Treadwell, and C. E. Schwartz. "Associations between Health Status and Utilities Implications for Policy." *Med Care* 37, no. 5 (1999): 479-89.

Llewellyn-Thomas, Hilary A., Elaine C. Thiel, and M. June McGreal. "Cancer Patients' Evaluations of Their Current Health States: The Influences of Expectations, Comparisons, Actual Health Status, and Mood." *Medical Decision Making* 12 (1992): 115-22.

Lubow, M., and L. Kuhr. "Pseudotumor Cerebri: Comments on Practical Management." In *Neuro-Ophthalmology*, edited by J.S. Glaser and J.L. Smith, 199-206. St. Louis: C.V. Mosby, 1976.

Lund-Johansen, M., F. Svendsen, and K. Wester. "Shunt Failures and Complications in Adults as Related to Shunt Type, Diagnosis, and the Experience of the Surgeon." *Neurosurgery* 35, no. 5 (1994): 839-44; discussion 44.

Maher, C. O., J. A. Garrity, and F. B. Meyer. "Refractory Idiopathic Intracranial Hypertension Treated with Stereotactically Planned Ventriculoperitoneal Shunt Placement." *Neurosurg Focus* 10, no. 2 (2001): E1.

McNeil, B, SG Pauker, and A Tversky. "On the Framing of Medical Decisions." In *Decision Making : Descriptive, Normative, and Prescriptive Interactions*, edited by David E. Bell, Howard Raiffa and Amos Tversky. Cambridge ; New York: Cambridge University Press, 1988.

Meyerowitz, B. E., and S. Chaiken. "The Effect of Message Framing on Breast Self-Examination Attitudes, Intentions, and Behavior." *J Pers Soc Psychol* 52, no. 3 (1987): 500-10.

Miceli, M., and C. Castelfranchi. "The Mind and the Future - the (Negative) Power of Expectations." *Theory & Psychology* 12, no. 3 (2002): 335-66.

Newborg, B. "Pseudotumor Cerebri Treated by Rice Reduction Diet." *Arch Intern Med* 133, no. 5 (1974): 802-7.

Powell, J. "Estimation of Semiparametric Models." In *Handbook of Econometrics*, edited by R. Engle and D. McFadden, 2444-521: North Holland, 1994.

Rasiel, E. B., K. P. Weinfurt, and K. A. Schulman. "Can Prospect Theory Explain Risk-Seeking Behavior by Terminally Ill Patients?" *Med Decis Making* 25, no. 6 (2005): 609-13.

Rothman, Alexander J., Peter Salovey, Carol Antone, Kelli Keough, and Chloe Drake Martin. "The Influence of Message Framing on Intentions to Perform Health Behaviors." *Journal of Experimental Social Psychology* 29 (1993): 408-33.

Rubin, R. C., E. S. Henderson, A. K. Ommaya, M. D. Walker, and D. P. Rall. "The Production of Cerebrospinal Fluid in Man and Its Modification by Acetazolamide." *J Neurosurg* 25, no. 4 (1966): 430-6.

Scheier, M. F., and C. S. Carver. "Dispositional Optimism and Physical Well-Being: The Influence of Generalized Outcome Expectancies on Health." *J Pers* 55, no. 2 (1987): 169-210.

Scheier, M. F., K. A. Matthews, J. F. Owens, G. J. Magovern, Sr., R. C. Lefebvre, R. A. Abbott, and C. S. Carver. "Dispositional Optimism and Recovery from Coronary Artery Bypass Surgery: The Beneficial Effects on Physical and Psychological Well-Being." *J Pers Soc Psychol* 57, no. 6 (1989): 1024-40.

Stratmann-Schoene, D., and T. Klose. "Health Values and Prospect Theory: A Comment." *Medical Decision Making* 21, no. 1 (2001): 57-59.

Sugerman, H. J., W. L. Felton, J. B. Salvant, A. Sismanis, and J. M. Kellum. "Effects of Surgically Induced Weight-Loss on Idiopathic Intracranial Hypertension in Morbid-Obesity." *Neurology* 45, no. 9 (1995): 1655-59.

Tanne, E., R. Rosenman, and D. Friesner. "Estimating the Costs of a Quiet Disease: Idiopathic Intracranial Hypertension." Working Paper, May 2008.

Tomsak, R.L., A.S. Niffenegger, and B.F. Remler. "Treatment of Pseudotumor Cerebri with Diamox (Acetazolamide)." *Journal of Clinical Neuro-Ophthalmology* 8 (1988): 93 - 98.

Treadwell, J. R., and L. A. Lenert. "Health Values and Prospect Theory." *Medical Decision Making* 19, no. 3 (1999): 344-52.

Tversky, A., and D. Kahneman. "Advances in Prospect-Theory - Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty* 5, no. 4 (1992): 297-323.

Von Neumann, John, and Oskar Morgenstern. *Theory of Games and Economic Behavior*. 60th anniversary ed. Princeton, N.J. ; Woodstock: Princeton University Press, 2004.

Wakker, P., and A. Tversky. "An Axiomatization of Cumulative Prospect-Theory." *Journal of Risk and Uncertainty* 7, no. 2 (1993): 147-75.

Wall, Michael. *Idiopathic Intracranial Hypertension (Pseudotumor Cerebri)* University of Iowa Health Care, 08-15-03 n.d. [cited 03-06-2008. Available from http://webeye.ophth.uiowa.edu/DEPT/IIH/pc_index.HTM.

Table 3.1
Data Descriptions

Better	Binary variable indicating patient's disease status. Equals 1 if patient perceived themselves as better off than they were prior to diagnosis.
Male	Binary sex variable, equals 1 if patient is male
White	Binary race variable, equals 1 if patient is white_non hispanic
Earnings	Real income values from the last full year the individual worked.
Working	Binary variable, equals 1 if patient was working the year prior to symptoms interfering with daily life
Health_Ins	Binary variable, equals 1 if patient has health insurance
MedDv	Binary variable indicating whether or not the patient has a background in health care
Vision	Binary variable, equals 1 if patient suffered from problems with vision
Obese	Binary variable, equals 1 if patient was obese at time of diagnosis
Headache	Binary variable, equals 1 if patient suffered from headaches at any time
Diet	Binary variable, equals 1 if patient attempted weight reduction
Invasive	Binary variable, equals 1 if patient underwent a surgical procedure, including subsequent lumbar punctures, but was not on medication for IIH.
Rx	Binary variable, equals 1 if patient was prescribed a medication for IIH other than acetazolamide, but did not receive acetazolamide, nor a surgical procedure
Aceta	Binary variable, equals 1 if patient was prescribed acetazolamide, but did not receive additional medication for IIH, nor a surgical procedure
RxInv	Binary variable, equals 1 if patient was prescribed a medication other than acetazolamide and underwent a surgical procedure, but did not receive acetazolamide.
AcetaRxInv	Binary variable, equals 1 if patient was prescribed a medication other than acetazolamide in addition to acetazolamide, and underwent a surgical procedure.
AcetaRx	Binary variable, equals 1 if patient was prescribed a medication other than acetazolamide in addition to acetazolamide, but did not undergo a surgical procedure.
AcetaInv	Binary variable, equals 1 if patient was prescribed acetazolamide and underwent a surgical procedure, but did not receive additional medication for IIH.

Table 3.2
Summary Statistics

	Sample				Registry			
Variable	**N**	**Mean**	**St. Deviation**	**Coeff. Of Variation**	**N**	**Mean**	**St. Deviation**	**Coeff. Of Variation**
Male[a]***	151	0.0662	0.2495	376.7493	732	0.1393	0.3465	248.6950
White[b]*	151	0.9205	0.2714	29.4799	732	0.8784	0.3270	37.2294
Earnings	151	6425.6000	5216.2400	81.1791	152	6410.3100	5202.3500	81.1560
Health_Ins	151	0.9205	0.2714	29.4799	194	0.9330	0.2507	26.8692
Working	151	0.8278	0.3788	45.7588	176	0.8239	0.3820	46.3697
MedDv	151	0.2252	0.4191	186.1214	196	0.2296	0.4216	183.6509
Vision[c]***	151	0.6689	0.4722	70.5939	732	0.5246	0.4997	95.2623
Obese[d]***	151	0.4901	0.5016	102.3463	732	0.1803	0.3847	213.3465
Headache[e]***	151	0.8477	0.3605	42.5306	732	0.2910	0.4545	156.2035
Diet***	151	0.4702	0.5008	106.5022	732	0.1708	0.3766	220.5140
Rx	151	0.0464	0.2110	455.0667	732	0.0191	0.1371	716.6301
Aceta**	151	0.0861	0.2814	326.8968	732	0.0355	0.1852	521.4498
Invasive***	151	0.2517	0.4354	173.0176	732	0.4194	0.4938	117.7395
AcetaRx	151	0.0464	0.2110	455.0667	732	0.0178	0.1322	744.1999
AcetaRxInv***	151	0.2119	0.4100	193.4823	732	0.0724	0.2593	358.1738
RxInv***	151	0.1656	0.3729	225.2465	732	0.0642	0.2453	382.0263
AcetaInv***	151	0.1391	0.3472	249.6347	732	0.0560	0.2301	410.8128

* implies Ho: Sample Mean = Registry Mean, rejected at significance level of 10%

** implies Ho: Sample Mean = Registry Mean, rejected at significance level of 5%

*** implies Ho: Sample Mean = Registry Mean, rejected at significance level of 1%

a) Incidence of IIH for men is approximately .3/100,000, compared to 1/100,000 women (Binder et al., 2004).

b) There is no evidence to suggest that race or ethnicity are significant determinants of IIH (Goodwin, 2006).

c) Approximately 20% - 68% of all patients with IIH experience vision problems (Binder et al., 2004).

d) Obesity is believed to be a risk factor for IIH, especially for women, with the incidence increasing from 1/100,000 to approximately 19/100,000 for obese females between the ages of 20 – 44 (Binder et al., 2004; IHRF, 2007b).

e) Approximately 90% of all patients with IIH experience headaches (Binder et al., 2004).

Table 3.3
MRE Results

	Dependant Variable = Better	
Parameter	**Estimate**	**95% Confidence Interval**
Intercept	------	---------
	------	---------
Male	-0.0881**	[-0.2250, -0.0250]
	(0.0651)	
White	-0.1244**	[-0.2099, -0.0079]
	(0.0595)	
Earnings	-5.1E-07**	[-6.82E-07, -5.1E-07]
	(2.0473)	
Health_Ins	0.7468***	[0.4382, 0.8684]
	(0.1354)	
Working	-0.0054**	[-0.0143, -0.0003]
	(0.0044)	
MedDv	0.0045*	[-0.0010, 0.0072]
	(0.025)	
Vision	-0.1259**	[-0.1749, -0.0273]
	(0.0440)	
Obese	0.1011*	[-0.0005, 0.1997]
	(0.0594)	
Headache	0.0634	[-0.0232, 0.1153]
	(0.0415)	
Diet	-0.1194**	[-0.2457, -0.0305]
	(0.0628)	
Rx	0.2063**	[0.0033, 0.2657]
	(0.0845)	
Aceta	0.4918***	[0.3398, 0.7892]
	(0.1311)	
Invasive	0.0598**	[0.0022, 0.0646]
	(0.0199)	
AcetaRx	0.2801**	[0.0098, 0.3324]
	(0.0989)	
AcetaRxInv	0.0250**	[0.0011, 0.0308]
	(0.0094)	
RxInv	0.0411**	[0.0014, 0.0619]
	(0.0206)	
AcetaInv	0.0766*	[-0.0104, 0.1037]
	(0.0346)	

Standard errors are in parentheses.

*Statistical Significance based on 90% confidence interval

**Statistical Significance based on 95% confidence interval

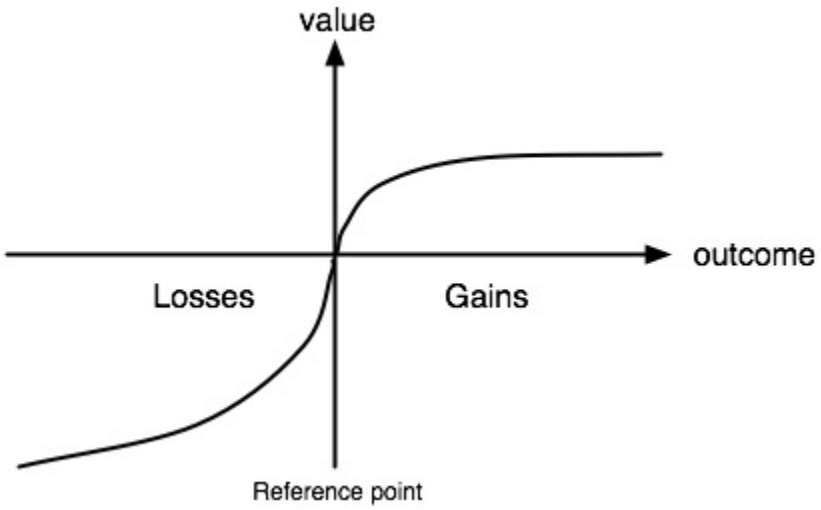
***Statistical Significance based on 99% confidence interval

Table 3.4

Hypothesis	p-value	Result	Expected Ordering	Actual Ordering
Rx > AcetaRx	<.0001	AcetaRx > Rx	Aceta	Aceta
Invasive > RxInv	<.0001	Invasive > RxInv	Rx	AcetaRx
Invasive > AcetaInv	<.0001	AcetaInv > Invasive	AcetaRx	Rx
Invasive > AcetaRxInv	<.0001	Invasive > AcetaRxInv	Invasive	AcetaInv
RxInv > AcetaRxInv	<.0001	RxInv > AcetaRxInv	RxInv	Invasive
AcetaInv > AcetaRxInv	<.0001	AcetaInv > AcetaRxInv	AcetaInv	RxInv
			AcetaRxInv	AcetaRxInv

Note: p-values are for Wilcoxon signed rank sum test of the null hypothesis that the effects of the two treatment paths on the dependent variable are equal.

Figure 3.1



Appendix A

Idiopathic Intracranial Hypertension

Individuals afflicted with intracranial hypertension suffer from elevated cerebro-spinal fluid (CSF) pressure in the skull. There are 2 types of intracranial hypertension (IHRF, 2007d). The first is *primary or idiopathic intracranial hypertension (IIH).* As the name implies, IIH arises spontaneously from an unknown cause. The other is *secondary intracranial hypertension,* which is associated with, and usually a side effect of, an identifiable cause such as a different disease, an intracranial blood clot, or certain drugs (IHRF, 2007a). The most common symptoms of increased intracranial pressure are headache and papilledema (Binder et al., 2004). The latter is particularly problematic due to the fact that over time swelling of the optic disc can lead to blindness or irreversible deterioration of vision (Giovannini and Chrousos, 2005).

Because the causes of *secondary intracranial hypertension* are known, its treatment can be tied to the patient's primary condition, and thus treated relatively effectively even though the prevalence of the disease is unknown. In this study we focus on IIH. IIH is most common among women of child-bearing age and occurs at an approximate rate of 1/100,000 (Binder et al., 2004); roughly 3 times that of males. Obesity is thought to be a risk factor for IIH, especially among women. The rate increases approximately 19 fold for females between the ages of 20 – 44 who are diagnosed as obese (Binder et al., 2004; IHRF, 2007b). However, while gender is a significant determinant of IIH, there is little evidence to suggest that race or ethnicity are significantly correlated with IIH (Goodwin, 2006).

There are several common treatments for idiopathic intracranial hypertension. Pharmaceutical treatments of IIH usually employ different types of diuretics, most commonly carbonic anhydrase inhibitors which reduce the production of CSF. Medications of this type are

the only ones known to be effective (Binder et al., 2004). Acetazolamide (originally sold under the trade name Diamox) is the most common medication of this sort, and the primary drug used to treat IIH. Its success in treating IIH has been well documented (Gücer and Viernstein, 1978; Lubow and Kuhr, 1976; Rubin et al., 1966; Tomsak, Niffenegger, and Remler, 1988). Another diuretic that has been shown to lower intracranial pressure is furosemide (Lasix). However, this drug does not appear to be as effective as acetazolamide, and in most cases is prescribed to individuals who cannot tolerate the latter (Binder et al., 2004; Gans, 2005). Other medicinal options do exist; however, little, if any, clinical evidence exists to support their efficacy (Friedman, 2005).

Due to the link between weight gain and IIH, physicians often recommend weight loss programs as a form of treatment. Previous studies have shown that weight loss appears to be an effective treatment for IIH (Johnson et al., 1998; Kupersmith et al., 1998). However, the importance of weight loss in this context remains unclear (Ball and Clarke, 2006; Binder et al., 2004).

Sugerman et al. (1995) report systematic improvement in patient symptoms following gastric bypass surgery[12]. Newborg (1974) documents resolution of papilledema in 9 patients after being treated with a diet alone. However, the small sample size prevents these results from being generalized to the entire population afflicted with IIH.

Surgical processes are generally reserved for patients who do not respond well to medicinal treatments. There are 2 primary types of surgery that can be performed. The first is optic nerve fenestration, where an incision is made in the sheath surrounding the optic nerve to relieve papilledema. The second involves the use of neurosurgical shunts, which are used to

[12] All of the patients in that study were on acetazolamide at one time, and it is unclear whether they continued to take the drug after surgery.

drain the CSF into another area of the body (IHRF, 2007c). The principle types of shunting procedures used to treat IIH are lumboperitoneal (LP) and ventriculoperitoneal (VP), although LP shunts are used most often as they are easier to insert (Binder et al., 2004; Friedman and Jacobson, 2004). Revisions are quite common with both procedures. LP shunts have a revision rate somewhere between 38% and 64% (Friedman and Jacobson, 2004). VP shunts appear to have a slightly lower revision rate in the range of 23% to 41% (Bynke et al., 2004; Lund-Johansen, Svendsen, and Wester, 1994; Maher, Garrity, and Meyer, 2001).

Repeated lumbar punctures are sometimes used as a surgical alternative. However, according to Binder et al. (2004), this is a less than ideal approach to treating IIH and should only be used as an emergency measure for patients who experience a sudden loss of vision resulting from serious cases of papilledema. Curry, Butler, and Barker (2005) maintain that the best surgical procedure for IIH remains unknown.

Appendix B

Monotone Rank Estimator (MRE)

The following model is based on Abrevaya and Hausman (1999), and is an extension of Han's (1987) *generalized regression model.* The latent dependent variable is as follows:

$$R^* = g(x\beta_0, \varepsilon), \tag{8}$$

where ε is an *i.i.d.* error disturbance, and *g* is an unknown function containing strictly positive partial derivatives at every point. The distribution of *R* then has the following c.d.f:

$$F_{R|R^*}(n \mid d) = \Pr(R \le n \mid R^* = d), \tag{9}$$

where *n* and *d* represent potential values for the dependent variable. For a model with a binary dependent variable, the probabilities of misclassification are:

$$\alpha_0 \equiv \Pr(R = 1 \mid R^* < 0) \tag{10}$$

$$\alpha_1 \equiv \Pr(R = 0 \mid R^* > 0). \tag{11}$$

The conditional c.d.f. becomes

$$F_R(n \mid R^*, d < 0) = \begin{Bmatrix} 0 & if & n < 0 \\ 1-\alpha_0 & if & n \in [0,1) \\ 1 & if & n \ge 1 \end{Bmatrix} \tag{12}$$

$$F_R(n \mid R^*, d \ge 0) = \begin{Bmatrix} 0 & if & n < 0 \\ \alpha_1 & if & n \in [0,1) \\ 1 & if & n \ge 1 \end{Bmatrix}. \tag{13}$$

To estimate the parameters we use the MRE, which is a rank estimator for semiparametric monotonic linear index models. The MRE consists of a vector $\hat{\beta}^{MRE}$ that maximizes the following objective function:

$$S^{MRE}(b) = \sum_i M(R_i) \cdot Rank(x_i b) \tag{14}$$

over the set $B \equiv \{b \in \Re^l : |b_l| = 1\}$, where $\Re$ represents the real line, M is an increasing function in *R*, $X'\beta$ is the linear index, *l* represents the number of covariates in *x*, and $|b_l|$ is the determinant of the *b* vector. Two comments are in order here. First, note that because the MRE is based on a rank-order process, there is no need to explicitly include an intercept in $x_i b$. Second, equations (12) and (13) imply that the stochastic-dominance conditions are fulfilled when $(1-\alpha_0) > \alpha_1$, which if it holds implies consistency of the parameter estimates.

The *Rank* function is defined by:

$$x_{i1}b < x_{i2}b < ... < x_{in}b \Rightarrow Rank(x_{im}b) = m. \quad (15)$$

Some examples of functions for *M* are given by Cavanagh and Sherman (1998). For robustness, $M(R) = Rank(R)$, for efficiency $M(R) = R$, or an intermediate alternative would be $M(R) = a\{R < a\} + R\{a \le R \le b\} + b\{R > b\}$, such that *a* and *b* are real numbers and $a < b$. By using a semiparametric approach we may be sacrificing some efficiency relative to a correctly specified parametric model (Powell, 1994); therefore, we used the second option to increase the efficiency of our estimates. Finally, the primary condition for consistency is that $\mathrm{E}[M(R)|X]$ is a nonconstant increasing function of $X'\beta$; however, a sufficient condition for consistency is that the distribution of *R* for a higher *R** first order stochastically dominates that of an *R* associated with a lower *R**.

CHAPTER FOUR

THE LATENT CHOICE MULTINOMIAL LOGIT MODEL: A SIMULATION

Abstract

Missing and ambiguous responses are often a major concern for applied researchers. The reallocation of those responses could allow for correction of potential measurement error and sample selection bias, giving the researcher efficient estimates. Using the "latent-choice multinomial logit model" (LCMNL), it is possible to determine whether these incomplete responses are more likely to belong to an outcome other than the one they are in. Simulations of the LCMNL are performed to determine whether the conditional probabilities estimated by the model are accurate enough to assess the likelihood that any given observation belongs to a particular outcome, and whether doing so improves parametric estimation. The data are created via a specified data generating process, allowing the true values of the incomplete responses to be known beforehand. Previous applications have only used real-world data. Tests imply that the reclassifications based on the estimated conditional probabilities are accurate, but the total number of reclassifications should be viewed with caution, as many of the ambiguous observations are not reassigned to one of the initial outcomes. The best method for dealing with ambiguous observations in empirical analysis is also assessed. The results indicate that the best course of action for managing the incomplete responses depends on the source of the ambiguity, i.e. whether these responses arise because of random events or true uncertainty.

Keywords: Latent choice multinomial logit, Don't know, Reclassification.

Introduction

Missing and ambiguous information in a dataset are a major concern for applied researchers of many fields. Estimation problems may arise due to this deficiency of information, leading to inaccurate results, and possibly limiting the types of analyses available to the researcher. The question of how best to treat the *don't know* (DK) response in contingent valuation data has received quite a bit of attention (Alberini, Boyle, and Welsh, 2003; Carson et al., 1998; Groothuis and Whitehead, 2002; Haener and Adamowicz, 1998; Wang, 1997). Previous methods of dealing with DK responses include: grouping the responses with the *no* outcome, which results in more conservative estimates (Carson et al., 1998); exclusion of the responses based on the assumption that they either imply indifference (Johannesson et al., 1993), or that they are not significantly different from the rest of the sample (Wang, 1997); and treating them as a middle response between *yes* and *no* in ordered categorical models based on the assumption that the respondents were truly uncertain (Groothuis and Whitehead, 2002; Wang, 1997). Wang (1997) provides four reasons why a respondent may provide an ambiguous response such as DK: 1) some individuals may not be willing to consider the setting of the question, thereby refusing to attempt to determine their preferences; 2) some may know their preferences, but for some reason answer ambiguously; 3) some make an attempt to examine their preferences and answer honestly; and 4) some may not make enough of an effort to determine their preferences, but are still answering truthfully at the time. Caudill and Groothuis (2005) argue that the first, second, and fourth explanations suggest that some of the DK responses may more appropriately be assigned to the *yes* or *no* outcome.

Based on the assumption that some ambiguous responses are misclassified Caudill (2006) extends Dempster, Laird, and Rubin's (1977) work on the expectations maximization (EM)

algorithm, developing a "latent-choice multinomial logit model" (LCMNL), which allows researchers to test for hidden alternatives while estimating a multinomial logit model with more options than are available given the data. This method of estimation also provides the researcher with an approximation of how many of the misclassified responses belong to each subcategory. Examples include: Caudill (2006) testing for hidden unemployment, Caudill, Ayuso, and Guillen (2005) testing for dishonest insurance claims, and Caudill and Groothuis (2005) testing whether DK responses in the contingent valuation literature are more like *yes* or *no* responses.

Using the conditional probabilities generated for each ambiguous observation by the LCMNL, it is possible to view the probability that a given response is misclassified. Caudill, Ayuso, and Guillen (2005), present an example of this when they review the cases of two individuals whose insurance claims are likely misclassified as either "honest," or "fraudulent." Caudill and Groothuis (2005) recalculate willingness to accept (WTA) and willingness to pay (WTP) estimates after reclassifying some DK responses to either *yes* or *no*, based on the estimated conditional probabilities. Our paper explores the possibility that inserting the reclassified ambiguous observations into the dependent variable could also serve as a method of controlling for the sample selection bias or measurement error that may arise due to these observations being absent from the sample, while simultaneously increasing the sample size. However, none of these applications directly assess the accuracy of the LCMNL, most specifically, how accurate the reassignments are and how this influences the statistical analysis of the data, as its only applications have been on real-world data without knowledge of the underlying data generating process.

The purpose of this paper is to perform simulations of the LCMNL to determine: 1) if the conditional probabilities generated by the model are accurate enough to assess the likelihood that

any given observation belongs to a particular outcome as Caudill, Ayuso, and Guillen (2005) and Caudill and Groothuis (2005) did, and 2) the implications for statistical analysis of data with ambiguous observations among dependent binary outcomes. Through the use of a specified data generating process, a random data matrix is created, inter-variable correlations from a real-world dataset are imposed on it, and a dependent variable is generated using known parameter values. As a result, the true values of the incomplete responses are known beforehand, which allows us to accomplish our first goal. Knowledge of the true parameter values allows us to accomplish the second objective by being able to test estimated coefficients for the data generating process against their true values.

The remainder of this paper proceeds as follows. First, the accuracy of the LCMNL in reassigning ambiguous responses via the conditional probabilities generated by the model is analyzed. Using the data generating procedure, three datasets with continuous dependent variables are created. The three datasets have equal population means, but different standard deviations in order to assess the performance of the LCMNL under different data conditions. Next, the continuous dependent variables are transformed into binary variables by choosing a cutoff point. A third outcome to represent ambiguous responses, e.g. the *don't knows*, is then artificially imposed on the three sets of generated data.

In the end two new dependent variables are generated for each of the three datasets by using two different methods to create the third (ambiguous) outcome. One method is designed to capture the measurement error often observed with survey data, i.e. the new ambiguous responses are randomly chosen and thus completely unrelated to the underlying continuous dependent variable. The other method is designed to represent respondent uncertainty in that the new responses are structured such that they fall very near the point where the underlying

continuous dependent variable are initially censored to create the binary variables. The conditional probabilities generated by the LCMNL are then used to reclassify these responses, which in turn are compared to their true values in order to assess the accuracy of the reclassifications. Tests imply that the reclassifications based on the estimated conditional probabilities are accurate, but the total number of reclassifications should be viewed with caution, as many of the ambiguous observations are not reassigned to one of the initial outcomes.

Analyses are also performed on each of the three datasets to determine the best method for handling incomplete responses in empirical analysis. A binary regression is initially performed on the original data prior to creating the ambiguous observations. Using the same dataset, the ambiguous responses from each of the new dependent variables are removed and binary regressions are run on the remaining observations. The estimates are then tested against their true values to determine whether a bias is introduced by removing these observations. The LCMNL is subsequently used to reclassify the ambiguous observations that had been removed for the second regression. The regressions are rerun for the each of the three datasets with the reclassified observations. Because the reassignments indicated by the LCMNL are not completely accurate, the Monotone Rank Estimator (Cavanagh and Sherman, 1998), an econometric technique used to adjust for misclassification of the dependent variable (Abrevaya and Hausman, 1999), is also applied to the reclassified data to see if it can aid in the correction of the estimates. Finally, ordered logistic regressions are performed on the data, both before and after reclassification, to determine the effect of directly incorporating the ambiguous observations into the analysis. Taken together, the results from the battery of comparisons described above indicate that the best course of action for managing the incomplete responses depends on the source of the ambiguity, i.e. whether these responses arise because of random

events or true uncertainty. Subsequent sections describe the LCMNL, the data generating procedure, empirical methodology, and results. The paper concludes with a summary of the findings and suggestions for future research.

The Model

Consider first the traditional multinomial logit model (MNL). The probabilities in this model are as follows:

$$P_{ij} = \frac{e^{\alpha_j + X_i\beta_j}}{\sum_{j=1}^{J} e^{\alpha_j + X_i\beta_j}} \quad \text{for } j = 1, \ldots, J, \tag{1}$$

where P_{ij} represents the probability that individual *i* chooses alternative *j*, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are parameters to estimate, and $\boldsymbol{X_i}$ is a vector of exogenous variables specific to individual *i*. The log-likelihood function is given by:

$$\ln L = \sum_{i=1}^{n} \sum_{j=1}^{J} d_{ij} \ln P_{ij}, \tag{2}$$

where *n* is the sample size and d_{ij} is a binary variable indicating whether individual *i* chose alternative *j*, for a total of J dummy variables.

The first step in developing the LCMNL is to extend the MNL to accommodate the case where the data contain missing information with respect to some individuals' choices. Caudill (2006) uses terminology generally reserved for nested logit models to help explain the modeling of hidden alternatives. The observable choices are referred to as branches. In the context of the application to DK responses (Caudill and Groothuis, 2005), assume that the DK branch contains three unobservable choices: *yes, no,* and DK. These are referred to as stems. As illustrated in Figure 1, the resulting logit has five final alternatives, two observed, and three unobserved.

By re-formulating the MNL, the probabilities for the LCMNL differ from those of the MNL. Assuming the branch associated with the DK responses is the third, the stem probabilities are now characterized as follows:

$$P_{i,3j} = \frac{e^{\alpha_{3j}+X_i\beta_{3j}}}{e^{\alpha_1+X_i\beta_1}+e^{\alpha_2+X_i\beta_2}+e^{\alpha_{31}+X_i\beta_{31}}+e^{\alpha_{32}+X_i\beta_{32}}+e^{\alpha_{33}+X_i\beta_{33}}}, \text{ for } j = 1,\ldots,3. \quad (3)$$

The probabilities of the remaining branches are then denoted by:

$$P_{ij} = \frac{e^{\alpha_j+X_i\beta_j}}{e^{\alpha_1+X_i\beta_1}+e^{\alpha_2+X_i\beta_2}+e^{\alpha_{31}+X_i\beta_{31}}+e^{\alpha_{32}+X_i\beta_{32}}+e^{\alpha_{33}+X_i\beta_{33}}}, \text{ for } j = 1,2. \quad (4)$$

Again, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the parameters that will be estimated and $\boldsymbol{X}_i$ is a vector of exogenous variables specific to individual i.[13] When the DK alternatives are hidden, the log-likelihood function is:

$$\ln L = \sum_{i=1}^{n} (d_{i1}\ln(P_{i1}) + d_{i2}\ln(P_{i2}) + d_{i3}\ln(P_{i,31}+P_{i,32}+P_{i,33})), \quad (5)$$

where d_{ij} remains a binary variable indicating whether individual i chose alternative j. If, however, the characterizations of the DK are known we can directly estimate the parameters of a new model, as follows:

$$\ln L = \sum_{i=1}^{n} (d_{i1}\ln(P_{i1}) + d_{i2}\ln(P_{i2}) + d^*_{i,31}\ln(P_{i,31}) + d^*_{i,32}\ln(P_{i,32}) + d^*_{i,33}\ln(P_{i,33})), \quad (6)$$

where $d^*_{i,3j}$ is an unobserved binary variable indicating whether observation i is associated with stem j. If the stem associations of the DK responses were observable, maximum likelihood could be used to directly estimate the parameters, as it would simply be a traditional multinomial logit model. But since these stem associations are not observable, setting the problem up in this manner allows for the maximum likelihood estimation of the parameters via the EM algorithm.

[13] As with any logit model, one of the $\boldsymbol{\beta}$ vectors must be set equal to 0 for identification purposes.

In the first stage of the algorithm, the expectations (E) step, the $d^*_{i,3j}$ s in equation (6) are replaced with their conditional probabilities, which are calculated as:

$$E(d^*_{i,3j} \mid d_{i3} = 1) = \frac{e^{\alpha_{3j} + X_i\beta_{3j}}}{e^{\alpha_{31} + X_i\beta_{31}} + e^{\alpha_{32} + X_i\beta_{32}} + e^{\alpha_{33} + X_i\beta_{33}}}, \text{ for } j = 1,\ldots 3. \quad (7)$$

In the context of the DK example, the conditional probabilities are interpreted as the probability that a response is a *yes, no,* or DK, given the actual response of DK.

During the second stage of the algorithm, the maximization (M) step, the log-likelihood in equation (6) is maximized, providing new estimates of the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The "E" and "M" steps are then repeated until the likelihood function converges to a maximum, at which point the standard errors can be calculated.

In order to apply the model to the case of misclassified responses, constraints must be imposed.[14] The constraints, $\boldsymbol{\beta}_{3j} = \boldsymbol{\beta}_j$ for $j = 1,2$, come from the Cramer and Ridder (1991) pooling condition for alternatives in a multinomial logit model.[15] The effect of the constraints is to pool observations from a particular stem with those from the corresponding branch. In the DK example, this would result in some of the ambiguous responses being aligned with *yes* or *no,* and some remaining DK.

Once the constraints have been imposed, the focus shifts to the intercepts of the model. Profile likelihood confidence intervals are created to test the hypothesis that $\boldsymbol{\alpha}_{3j} \to -\infty$ for $j = 1$, 2. Failure to reject this hypothesis implies that a statistically significant number of observations associated with stem j can not be pooled with those from branch j. For example, the test could indicate that the number of reassignments from DK to either *yes* or *no* is not statistically

[14] The unconstrained version of the model faces identification problems. See Caudill (2006) for a more detailed explanation.
[15] The constraints allow for maximum likelihood estimation via algorithms other than the EM (Caudill and Groothuis, 2005).

significantly different from zero. Profile likelihood confidence intervals are used instead of classic confidence intervals because the latter will always reject the null hypothesis. Classic confidence intervals assume that the sampling distribution of the MLE is asymptotically normal, forcing the boundary values to be symmetric in relation to the estimate. As a result, the confidence interval will always have a lower bound and the hypothesis will always be rejected. The boundary values of the profile likelihood confidence interval are not restricted in this manner because it is based on the chi-square distribution of the likelihood ratio statistic.

Caudill (2006) presents the following discussion of profile likelihood confidence intervals, which in turn draws from Venzon and Moolgavkar (1988). Let $\boldsymbol{\theta}$ be a parameter vector of length *k*, and $\boldsymbol{\theta}^*$ be the corresponding maximum likelihood estimate (MLE). Furthermore, let $L(\boldsymbol{\theta})$ be the log-likelihood function for this parameter vector, where $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset R^n$. The profile likelihood technique maximizes the log-likelihood function, treating all parameters but one, say β, as nuisance parameters, which are allowed to vary. This results in:

$$\Theta_j(\beta) = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : \theta_j = \beta\}, \tag{8}$$

with a profile likelihood function of:

$$L_j(\beta) = \max L(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_j(\beta), \tag{9}$$

and a confidence interval for β_j of :

$$\{\beta : L(\boldsymbol{\theta}^*) - L_j(\beta) \leq 0.5 q_1(1-\alpha)\}, \tag{10}$$

where $q_1(1-\alpha)$ represents the $(1-\alpha)^{th}$ percentile of the chi-square distribution given one degree of freedom.

In the case of the DK application of the LCMNL, the β_j represents the intercept terms associated with the *yes* and *no* stems, implying that the profile likelihood confidence interval from equation (10) must be constructed for α_{31} and α_{32}. To construct the $(1-\alpha)\%$ confidence interval, the likelihood function is maximized after fixing the alpha parameter, and allowing the remaining nuisance parameters to be estimated. This process is repeated, increasing or decreasing the intercept of interest and comparing it to the original MLE until parameters are discovered that leave $\alpha/2$ in the tail of the chi-square distribution of the likelihood ratio test statistic. If the confidence interval for the given stem intercept term contains a lower bound, then it is possible for a significant number of observations associated with that stem to be pooled with its respective branch.

Data Generating Procedure

The LCMNL is tested on data created via a specified data generating process, a flowchart of which can be viewed in Figure 2. An important consideration when generating the sample data is that it could be considered to have been drawn from a population with multiple defining characteristics, $\boldsymbol{X}_i$. Kaiser and Dickman (1962) propose a procedure to impose inter-variable correlations from a real world dataset onto one that is randomly generated. To begin, a correlation matrix is calculated for a desired population. Principal component analysis (PCA) is then performed on the correlation matrix in order to capture the inter-variable correlation pattern. If the sample data from which the correlation matrix is derived are composed entirely of binary variables, it is often recommended that tetrachoric correlations be used when performing PCA. The tetrachoric correlation coefficient between two variables is an estimate of what the Pearson

correlation would be if the variables were continuous.[16] After performing the principal component analysis, Kaiser and Dickman (1962) create a data matrix, $\hat{\boldsymbol{Z}}$, containing k uncorrelated random normal variables with a specified mean and standard deviation, and N observations. Once the data matrix is in hand, the sample score matrix is calculated as follows:

$$\hat{\boldsymbol{X}}_{(Nxk)} = \hat{\boldsymbol{Z}}_{(Nxk)} \boldsymbol{P}_{(kxk)}, \tag{11}$$

where $\boldsymbol{P}$ is the matrix made up of the principal components, and $\hat{\boldsymbol{X}}$ is the new data matrix with the inter-variable correlations imposed.

The next step in the data generation process is the creation of the dependent variable. This is accomplished using the following equation:

$$\boldsymbol{Y}^{*}_{(Nx1)} = \hat{\boldsymbol{X}}_{(Nxk)} \boldsymbol{B}_{(kx1)} + \boldsymbol{E}_{(Nx1)}, \tag{12}$$

where $\hat{\boldsymbol{X}}$ is the data matrix from (11), $\boldsymbol{B}$ is a vector of specified parameter values, $\boldsymbol{E}$ is a vector of stochastic error terms, and $\boldsymbol{Y^*}$ is a continuous variable, which will later be transformed to a categorical variable. Because the LCMNL is a logistic model, the error terms are randomly generated from a logistic distribution.

To provide some real-world context, and in light of a future application, the correlation matrix for the PCA is derived from a dataset of patients suffering from idiopathic intracranial hypertension.[17] A survey of patients with this disorder resulted in a significant number of blank responses to whether treatment had improved their conditions. The ability to reassign some of these incomplete observations to the *better* or *not-better* outcome could allow for more accurate estimates of how various treatments and other factors influence patient well-being. Also,

[16] The tetrachoric correlations can be computed in SAS using the POLYCHOR macro, which can be found at: http://ftp.sas.com/techsup/download/stat/polychor.html

[17] The data come from the Intracranial Hypertension Registry: http://www.ihrfoundation.org/news/registry.asp. This is a project co-sponsored by the Intracranial Hypertension Research Foundation of Vancouver, Washington and the Casey Eye Institute at the Oregon Health and Science University (OHSU).

excluding ambiguous observations is generally discouraged on account of reduced sample size and econometric efficiency, as well as potential sample selection bias (Groothuis and Whitehead, 2002; Wang, 1997). Therefore, nine variables were chosen for the PCA from a highly significant binary logistic model estimating the effects of various treatments and patient characteristics on an individual's post-treatment disease status.[18]

Three sets of $\hat{\boldsymbol{X}}$s are created with the data generating procedure just described, each containing nine variables and 500 observations. The variables in the first $\hat{\boldsymbol{X}}$ are constructed to have a population mean of zero and standard deviation of one (s=1). The population means of the variables in the second and third $\hat{\boldsymbol{X}}$s are equal to the original sample, but have standard deviations of s=2 and s=3 respectively, in order to assess the performance of the LCMNL under different data conditions.

The next step in transforming the data for use in the LCMNL test is to censor the dependent variable from equation (12). Because ***Y**** is continuous, it represents the latent log odds ratio in a logistic context. Take, for example, the case where a patient is asked to claim a post-treatment disease status of "better" or "not-better". If $p_i = \Pr(better) = \frac{1}{1+e^{-(\alpha_i + x_i b)}}$, then $y_i^* = \ln[p_i/(1-p_i)] = \alpha_i + x_i b$, which implies that at $y_i^* = 0$, $p_i = (1-p_i)$. Therefore, the ***Y**** variable in each new dataset is transformed such that $y_i = 2$ if $y_i^* > 0$, and $y_i = 1$ otherwise. The observations with the value one will be referred to as belonging to outcome one, and those with the value two as belonging to outcome two.

To test the accuracy of the LCMNL reclassifications as discussed in the introduction of the paper, a third outcome to represent the ambiguous responses is created. Two new dependent

[18] The parameter values for equation (12) were also chosen from this regression.

variables are generated for each dataset through the use of two different methods of constructing this outcome. First, 20% of the observations are randomly chosen and the dependent variable for these observations is set to the value 3. The observations in this new outcome represent the measurement error that is often observed with survey data in that the missing data is completely unrelated to ***Y****. For example, data could be missing because of random errors in record keeping or transmission, or because respondents answered ambiguously for some reason other than being truly unsure of their preferences.[19] Second, observations are constructed to represent respondent uncertainty, i.e., those for which the respondents might reasonably truly not know or are ambivalent or effectively indifferent about how to answer. To do so a parameter δ is arbitrarily chosen such that:

$$\begin{aligned} &y_i = 1 \, if: \; y_i^* < (0 - \delta) \\ &y_i = 3 \, if: (0 - \delta) \le y_i^* \le (0 + \delta) \\ &y_i = 2 \, if: \; y_i^* > (0 + \delta). \end{aligned} \qquad (13)$$

In this paper $\delta = 1$ to ensure that there are a reasonable number of ambiguous observations representing respondent uncertainty in each dataset. Summary statistics for the new datasets can be viewed in Table 1.

Empirical Methodology

The purpose of this paper is to assess the accuracy of the conditional probabilities used to determine whether a given observation is misclassified, and to determine the best course of action for dealing with incomplete responses in empirical analysis. Tests used to analyze the LCMNL in previous applications are also presented here in order to compare the model's performance in this paper to that of other studies.

[19] See page 73 for a list of reasons why an individual may provide an ambiguous response.

As suggested by Caudill (2006), two of Cramer and Ridder's (1991) pooling tests are run on each set of ambiguous responses from each $\hat{\boldsymbol{X}}$. This test consists of a likelihood ratio test of the hypothesis that combining all of the ambiguous (i.e. outcome=3) responses with those from outcome=1, for example, and running a binary logistic regression on outcomes one and two, is statistically equivalent to running the multinomial logistic regression with all three outcomes. The test is then repeated, this time combining all outcome=3 observations with those from outcome=2. The parameter estimates are indistinguishable: $H_0 : \beta_3 = \beta_j$, where j = 1,2. Rejection of the null hypothesis implies that not all of the outcome=3 responses can be pooled with those from outcome=j, where j=1,2. If the two tests performed on a given dataset agree, i.e. they don't indicate that all responses could be grouped with outcome=1 for example, and not outcome=2, the implication is that it may be possible for some of the observations to belong to one group and some to the other.

The LCMNL does not physically reclassify individual observations. Instead, the approximate number of responses that are reassigned within the procedure is calculated by summing the conditional probabilities from equation (7) for each stem of the dependent variable, i.e. $\sum \hat{d}^*_{i,3j}$, where j=1,2,3. At this point, the profile likelihood confidence intervals are used to determine whether a statistically significant number of reassignments occurred. If the number of reassignments is significantly greater than zero, this indicates that the variable in question does contain hidden alternatives.

Previous applications of the LCMNL have suggested that the estimates are quite similar to those from the MNL, i.e. one that does not allow for hidden alternatives. Therefore, the estimates from the LCMNL are compared to those from the MNL. A Wilcoxon signed rank sum test, the nonparametric alternative to the paired t-test, is performed on the predicted probabilities

from the two models. This test consists of ranking the absolute differences between each pair of predicted probabilities and calculating the Wilcoxon signed rank statistic to test the null hypothesis that the median difference is zero. If the MNL restriction is valid, the sum of the ranks for the positive differences should approximately equal the sum of the ranks of the negative differences. Recall that in the LCMNL the coefficients are constrained such that $\beta_{3j} = \beta_j$, where j = 1,2. Therefore, the null hypothesis is that the median predicted probabilities are equal, i.e. $\hat{p}_i(y = j)_{LCMNL} = \hat{p}_i(y = j)_{MNL}$, where j=1,2,3. If the null hypothesis is rejected the LCMNL and the MNL will both be compared to the true data to determine which is better.

To test the LCMNL against the MNL the predicted probabilities from each model are compared to those from the data generating process allowing us to determine the number of observations that each model predicts correctly. A two by two contingency table is constructed using the number of correct and incorrect predictions from each model, and a chi-square test of homogeneity is performed. If the null hypothesis that the models perform equally well is rejected, the counts in the interior of the table will indicate which is better. Assuming the LCMNL can reassign the ambiguous responses better than by chance, it should outperform the MNL.

Though the LCMNL does not physically reclassify individual observations, it is possible to determine which outcome each ambiguous response most likely belongs to by examining the conditional probabilities generated for each one (see equation 7). For example, given that an observation is classified as a DK, one can discern the probability that it is actually a *yes,* a *no,* or truly a DK. The incomplete responses are then physically reassigned based on their largest conditional probability. For example, if the LCMNL predicts the probabilities for a given outcome=3 response to be .6 for outcome one, .3 for two, and .1 for three, the response is

reclassified as belonging to outcome one (Caudill and Groothuis, 2005). To analyze the accuracy of the reassignments, the reclassified observations are compared to their known values, and t-tests are performed to determine whether the percentages of accurately reassigned observations are significantly different than 50%. That is, we test whether the reassignments indicated by the LCMNL are more accurate than those re-assigned through chance. Recall that the original binary dependent variable, ***Y***, is generated by censoring ***Y**** at the point on its distribution where $p_i = (1 - p_i)$.[20] Because the ambiguous observations representing respondent uncertainty are created such that they are within δ of the ***Y**** censoring point (see equation 13), on average they should have lower y_i^* values than the ambiguous observations representing measurement error, which are randomly chosen. As a result, the conditional probabilities estimated by the LCMNL (see equation 7) should be more accurate for the observations representing measurement error.

Ambiguous Responses in Empirical Analysis

The issue of how to manage ambiguous responses in survey data has received a great deal of attention. Therefore, regressions are performed to examine the best method of handling ambiguous responses in the empirical analysis. See Figure 3 for an organizational chart of this process. Recall that one method of dealing with ambiguous observations is exclusion. So the first question is whether removing these observations generates a bias in the estimates. The exclusion of the randomly created outcome=3 observations simulates the measurement error that is common in survey data, such as a response getting coded improperly. Removing the structurally created outcome=3 responses simulates sample selection bias, e.g. individuals who

[20] See discussion on page 83.

respond ambiguously for a reason. A binary regression is first run on the original data, i.e. prior to the creation of the third outcome, and the estimates are tested against the true parameters. Wald tests are used to test for jointly significant differences, and t-tests for individual differences. Regressions are then run on each of the dependent variables, minus the incomplete responses. By testing each of these parameter estimates against their true values, we are able to determine whether a bias is created when the ambiguous responses are absent, or if the only difference is a loss of efficiency.

Next the outcome=3 observations that were removed are physically reclassified according to the conditional probabilities generated by the LCMNL. Those that are reassigned to one of the original outcomes, i.e. outcome one or two, are inserted back into the binary dependent variables. The regressions are rerun and the estimates are retested against their true values to determine what effect including the probabilistically reclassified observations has on the bias if it exists. Assuming a bias is created by the removal of the ambiguous observations, inserting the reclassified observations back into the dependent variable is expected to help alleviate it if enough of the ambiguous observations are accurately reclassified to outcomes one and two.

One potential problem with the regressions performed on the reclassified data is that unless the reclassifications indicated by the LCMNL are completely accurate, the dependent variables are subject to misclassification error. Failure to control for this when estimating discrete-response models via traditional techniques such as logit or probit, can result in inconsistent estimates (Hausman, Abrevaya, and Scott-Morton, 1998). To control for the misclassification error Abrevaya and Hausman (1999) recommend semiparametric estimation, more specifically the *monotone rank estimator* (MRE) from Cavanagh and Sherman (1998), as an alternative to parametric estimation. Unlike the parametric approach, semiparametric

estimation does not require that the mismeasurement be modeled correctly in order to obtain consistent estimates. Therefore, the MRE is used to adjust for the potential misclassification bias.[21] The accuracy of these estimates is then tested.

First, the MRE is performed on the original data (prior to the creation of ambiguous responses) and confidence intervals used to test whether the estimates are significantly different than the normalized true parameter values. The estimates from the MRE are not directly comparable to the true parameters, because the MRE estimates the coefficient vector, which does not include an intercept term, up to scale such that the sum of their squares is equal to one. Therefore, the true parameters are normalized in a similar manner to obtain some comparability with the estimates from the MRE. Next, the MRE is performed on the binary dependent variables containing the reclassified ambiguous observations, and the estimates again tested against the normalized true parameters. Because the confidence interval tests provide only a rough estimate of the MRE's performance, a Wilcoxon signed rank sum test is performed on the $X_i\hat{\beta}$s that are common to both regressions to test for differences between the original known responses and the reclassified responses. The null hypothesis is that the median $X_i\hat{\beta}$s are equal, i.e. $X_i\hat{\beta}^{MRE}_{Original} = X_i\hat{\beta}^{MRE}_{Reclassified}$.

Finally, the incomplete responses are treated as an intermediate outcome corresponding to respondent uncertainty, and ordered logistic regressions are performed on the original datasets with ambiguous observations, as well as those where the ambiguous observations have been physically reclassified. The expectation is that the original data containing observations structured to represent respondent uncertainty will perform better, because these responses are designed such that they fall between the outcome one and two responses, and a review of

[21] See Appendix A for a detailed discussion of the MRE.

equation 13 reveals a data generating process equivalent to the assumed model structure underlying an ordered logistic regression and likelihood function.

Results

Though the primary foci of this paper are the accuracy of the reclassifications indicated by the conditional probabilities from the LCMNL and the handling of the ambiguous observations in empirical analysis, we first analyze the results from the LCMNL model itself, and compare them to previous applications in the literature. The pooling test of the null hypothesis that all ambiguous observations could be grouped with one of the other outcomes, i.e. $H_0 : \beta_3 = \beta_j$, where j = 1,2, is strongly rejected in each case at p<.01 or smaller.[22] These findings appear to contradict those of Carson et al. (1998), which support the method of grouping ambiguous responses with a different outcome, e.g. grouping the DK responses with the *no* responses in contingent valuation data. The LCMNL can be used to determine whether it is possible for some of the ambiguous observations to be pooled with outcome=1, and some with outcome=2.

In each of the previous studies, as in this paper, summing the conditional probabilities indicates that observations are likely to be reassigned to all outcomes, including the one of origin, e.g. the ambiguous outcome (Caudill, 2006; Caudill, Ayuso, and Guillen, 2005; Caudill and Groothuis, 2005). These results are presented in the top half of Table 2 for the ambiguous observations representing measurement error and those representing respondent uncertainty from the three $\hat{X}$ s, i.e. s=1,2,and3. The accuracy of the reassignments can not be compared to those

[22] Pooling tests are discussed in more detail on page 85.

from prior applications of the model in the literature because the true classifications were unknown in those studies.

The results from the profile likelihood confidence interval tests of the intercepts associated with the three - one (α_{31}) and three - two (α_{32}) stems are mixed.[23] For the dataset with an *s*=1 and observations structured to represent respondent uncertainty, both tests reject the null hypothesis that $\alpha_{3j} \to -\infty$, where *j*=1,2. This result implies that the number of ambiguous observations reassigned to the first and second outcomes is statistically significantly greater than zero[24]. Just the opposite is true for the two datasets with an *s*=2, in that none of the intercepts contain lower bounds, implying that a statistically significant number of ambiguous observations could not be pooled with either the outcome=1 or outcome=2 observations. This has not been observed in the literature, but there have only been two applications of the LCMNL in which these tests have been performed. In the three remaining datasets only α_{32} is constrained away from negative infinity, implying that a statistically significant number of ambiguous observations are reassigned to outcome=2, but not outcome=1. These findings are similar to those from the WTA and WTP applications of the LCMNL (Caudill and Groothuis, 2005), where only the intercepts associated with the DK-*no* responses contain lower bounds.

Next, the results from the LCMNL regressions on all of the dependent variables in this paper are compared to those from the corresponding MNL regressions. The results from the LCMNLs for the datasets with an initial standard deviation of two can be viewed in Table 3, while those for the corresponding MNL regressions can be viewed in Table 4.[25] Prior studies

[23] See equations 6 and 7 and discussion on page 79.
[24] This finding is similar to Caudill (2006), where the tests indicate that a significant number of "own-account self-employed" individuals are more like employers, and others more like the unemployed.
[25] The results for the dataset with an *s*=2 were arbitrarily chosen to be presented. The results for the remaining datasets are similar, and are available from the author upon request.

report that the estimates from the LCMNL and MNL are quite similar. This also appears to be the case here for the data containing observations designed to represent measurement error, but not necessarily for the data with observations representing respondent indifference. The Wilcoxon signed rank sum test of the hypothesis that the median predicted probabilities are equal, i.e. $\hat{p}_i(y=j)_{LCMNL} = \hat{p}_i(y=j)_{MNL}$, where j=1,2,3, is rejected at p=.04 and smaller in all but two of the tests, both from datasets with ambiguous observations representing measurement error.

The Chi-square test of homogeneity performed on the two by two contingency table of correctly and incorrectly predicted responses from the MNL and LCMNL is rejected in each instance at p<.0001. Four of the tests suggest that the LCMNL outperforms the MNL. The test of the data with an *s*=1 and observations representing measurement error indicates that the MNL is better than the LCMNL, and in the test of the data with an *s*=2 and observations representing respondent uncertainty the number of correctly predicted responses is equal across the models.

The physical reassignments based on the conditional probabilities can be viewed in Table 2. Similar to the findings of Caudill and Groothuis (2005), the number of within model reassignments mentioned above varies from the number of observations that are physically reclassified. For the dataset created with an s=1, the conditional probabilities of the 100 ambiguous responses representing measurement error indicate that 11 observations should be assigned to outcome=1, 52 to outcome=2, and 37 should remain as outcome=3. While the probabilities of the 116 observations representing respondent uncertainty from the data with an *s*=1 indicate that the assignments should be 22 to outcome=1, 49 to outcome=2, and 45 to outcome=3. With an *s*=2, the LCMNL applied to the data with observations representing measurement error (respondent uncertainty) indicates that of the 100 (55) ambiguous

observations 24 (2), 30 (19), and 46 (34) should be assigned to outcomes one, two, and three, respectively. For s=3, the assignments of the observations representing measurement error (respondent uncertainty) are 31 (11), 35 (22), and 34 (12), out of 100 (45) ambiguous responses, to outcomes one, two, and three, respectively.

We now examine the accuracy of the reassignments from group three to the other two groups. For the data with an s=1 and observations representing measurement error, 49 of the 63 (78%) reclassifications are correct. This is true for 51/54 (94%) and 62/66 (94%) of the ambiguous observations representing measurement error from the datasets with s=2 and s=3, respectively. One-tailed t-tests are performed to determine whether these percentages are significantly greater than 50%. The tests for the observations corresponding to measurement error are all significant at conventional levels of confidence, with $p<.0001$ or smaller. The percentages for the observations representing respondent uncertainty are somewhat lower: 42/71 (59%) for s=1, 15/21 (71%) for s=2, and 24/33 (73%) for s=3. The one-tailed t-tests indicate that these percentages are also significantly greater than 50% at conventional levels ($p=.06$ for s=1, $p=.02$ for s=2, and $p=.003$ for s=3). So in terms of overall accuracy, the LCMNL performs better on the ambiguous observations representing measurement error than it does on those corresponding to respondent uncertainty.

The preceding tests do not differentiate between reclassifications from outcome=3 to either outcome 1 or 2. The results, however, suggest that the reclassifications tend to be weighted heavily towards one outcome or another for certain datasets. We therefore consider the percentage of correct reassignments to outcomes 1 and 2, respectively. For the group three observations representing measurement error from the dataset with an s=1, of the 52 observations that are reassigned to the second outcome, 39 (75%) truly belong there. This is true of 29/30

(97%) for *s*=2, and 33/35 (94%) for *s*=3. The percentages for the first outcome are as follows: 10/11 (91%) for *s*=1, 22/24 (92%) for *s*=2, and 29/31 (94%) for *s*=3. All are significantly greater than 50% at better than the 1% level. Once again, the corresponding percentages for the observations representing respondent uncertainty appear to be slightly lower. For the second outcome they are: 27/49 (55%) for *s*=1, 13/19 (68%) for *s*=2, and 16/22 (73%) for *s*=3. The p-values for the one-tailed t-tests are: for *s*=1, p=.24; for *s*=2, p=.06; and for *s*=3, p=.01. The percentages for the first outcome are: 15/22 (68%), 2/2 (100%), and 8/11 (73%) for *s*=1, 2, and 3, respectively. The one-tailed p-values are p=.04 for *s*=1, and p=.07 for s=3.

As expected, the LCMNL performs better in terms of reclassifications when applied to the ambiguous observations corresponding to measurement error. This holds true even after breaking down the reclassifications by outcome. At the same time, all but one of the percentages of accurately reassigned observations, both overall and by outcome, are significantly greater than 50%. Therefore, it appears that the conditional probabilities produced by the LCMNL are accurate enough to analyze individual observations like Caudill, Ayuso, and Guillen (2005) and Caudill and Groothuis (2005) did, especially if there is reason to believe the incomplete responses are due to random events and not to random responses related to (approximate) respondent indifference.

Ambiguous Responses in Empirical Analysis

To determine the ideal method for handling the ambiguous observations in empirical analysis, regressions are first run on the original data, i.e. prior to the creation of the third outcomes. The results from these regressions are presented in Table 5. Wald and t-tests indicate that these estimates are not significantly different than the true parameter values. Next, regressions are run

on each of the dependent variables, minus the ambiguous responses, and the estimates again tested against their true values. Table 6 contains the results from these regressions. Only one of the coefficient estimates from the regressions on the data with measurement error is rejected, and only at the 10% level. This implies that bias is not a concern if the ambiguous observations occur as a result of random circumstances. Each regression on the data containing respondent uncertainty produces between one and five estimates that significantly vary from their true values at traditional levels, indicating that bias is a concern if the incomplete responses arise because of true uncertainty on the part of the respondents.

Inserting the physically reclassified observations back into the dependent variables appears to increase the bias in the regressions performed on the data with respondent uncertainty, and create it in the regressions performed on the data with measurement error. The results from these regressions are presented in Table 7. In each regression, the number of parameter estimates that are significantly different from their true value increases, or at best doesn't change. The question now is whether using the MRE, a procedure used to control for misclassification of the dependent variable, on the reclassified data can correct the distortion in the estimates.

Table 8 contains the results from the MRE regressions. Two methods are used to test the accuracy of the estimates from the MRE. The MRE is first performed on the original data, i.e. prior to the creation of the ambiguous outcomes, and the estimates are tested against the normalized true parameters via confidence intervals. As mentioned above, the estimates from the MRE are not directly comparable to the true parameters, but the MRE performs quite well in that only two estimates, both from the dataset with an s=3, differ significantly from the normalized true parameters at the one percent level.

Next, the MRE is performed on the dependent variables containing the reclassified observations, and the estimates are again tested against their true values. Five estimates, three from the dataset with an s=3, now show up as significantly different than the normalized true parameters at the one percent level, but overall the magnitudes of the estimates are similar to those from the MRE regressions performed on the original data. Because the confidence interval tests provide only a rough estimate of how the MRE performs, a better test may be to compare the vectors comprised of $X_i\hat{\beta}$s that are common to both regressions, i.e. the MRE regression on the original data and the one on the reclassified data. The $\boldsymbol{X\hat{\beta}}$ vectors are tested using a Wilcoxon signed rank sum test, the nonparametric alternative to the paired t-test. This test is repeated for each dataset. All of the tests fail to reject the null hypothesis that the medians of the $\boldsymbol{X\hat{\beta}}$ vectors are equal with p-values of .23 and higher. These findings indicate that the MRE is successful at correcting the bias created from the removal of the observations representing respondent uncertainty, and/or the insertion of either type of reclassified ambiguous observations.

The results from the ordered logistic regressions can be viewed in Table 9. The regressions run on the original data containing observations structured to represent respondent uncertainty perform very well. In fact, t-tests indicate that only one of the coefficient estimates differs significantly from its true value at the 5% level, and none differ at the 1% level. This was expected, as these responses are designed such that they fall between the outcome one and two responses (see equation 13). This supports Wang's (1997) finding that truly uncertain responses can be directly incorporated into the analysis via ordered categorical models. The ordered logit performs poorly when applied to the data with observations representing measurement error, which was also expected being as this type of ambiguous observation is randomly created, and

thus unrelated to the underlying Y^* variable. The ordered logit also performs relatively poorly on the datasets where the ambiguous observations have been physically reassigned based on their estimated conditional probabilities from the LCMNL.

Conclusions

The focus of this paper is twofold. First, simulations are performed to determine whether the conditional probabilities generated by the LCMNL are accurate enough to assess the likelihood that a given observation belongs to a particular outcome as Caudill, Ayuso, and Guillen (2005) and Caudill and Groothuis (2005) did. Second, we attempt to determine the implications for statistical analysis of data with ambiguous observations among dependent binary outcomes. In addition, we compare our results to those from previous applications of the LCMNL.

It appears that the reclassifications indicated by the LCMNL's estimated conditional probabilities are quite accurate, implying that this could be an effective tool for analyzing individual observations believed to be misclassified. However, the total number of reclassifications should be viewed with caution, as many of the ambiguous observations were not reclassified to one of the initial outcomes. This is true not only of the physical reassignments, but also of the approximate number indicated by the summation of the conditional probabilities. As expected, the reclassifications associated with the models using observations designed to represent measurement error are more accurate than those associated with models using observations representing respondent uncertainty. Recall that the original binary dependent variable, ***Y***, is generated by censoring ***Y**** at the point on its distribution where $p_i = (1 - p_i)$.[26] Since the ambiguous observations representing measurement error are chosen at random versus

[26] See discussion on page 83.

those representing respondent uncertainty, which are created such that they are within δ of the ***Y**** censoring point (see equation 13), on average the former have higher y_i^* values. As a result, the conditional probabilities estimated by the LCMNL (see equation 7) are more accurate for the observations representing measurement error.

The best method for handling incomplete responses in data analysis appears to depend on the source of the ambiguity. Results indicate that exclusion of the ambiguous responses only generates a bias in the estimates if the observations appear because of actual uncertainty on the part of the respondents. And though all of the percentages of accurate reclassifications are significantly greater than 50%, using the physically reassigned observations in the binary model only appears to worsen the amount of bias in the regressions performed on the data with observations representing respondent uncertainty, and create it in the regressions performed on the data with observations representing measurement error. Therefore, if there is reason to believe that the incomplete responses occur because of random events such as errors in record keeping or transmission, the best method appears to be excluding these observations. Because this will result in biased estimates if the ambiguous observations occur as a result of actual uncertainty, the MRE, a method of controlling for misclassification in the dependent variable, is run on the dependent variables after incorporating the physically reclassified observations. Tests indicate that the MRE successfully corrects the distorted estimates.

Given that the observations structured to represent respondent uncertainty lie in the middle of the ***Y**** distribution (see equation 13), ordered logistic regressions are also performed as an alternative form of estimation. As expected the regressions on the data with observations representing respondent uncertainty prior to reclassification perform very well in that none of the parameter estimates are significantly different from their true values at the 5% or 1% levels.

This implies that if there is reason to believe the incomplete responses are due to true uncertainty on the part of the respondents, they should be directly incorporated into the analysis via an ordered categorical model. If the researcher is uncertain of the reason for the incomplete responses, as may often be the case, the best course of action appears to be integrating the physically reclassified observations into the dependent variable, followed by the use of the MRE.

Although many of these findings cannot be directly compared to those from previous papers due either to the lack of information about the underlying data generating process, or the lack of comparable analyses, the results that can, appear to be quite similar. For example, in other applications of the LCMNL (Caudill, 2006; Caudill, Ayuso, and Guillen, 2005; Caudill and Groothuis, 2005), as in this paper, the model indicates that some ambiguous observations should be assigned to every outcome, including the one of origin, e.g. the DK outcome. Also, in most of the previous examples the reclassifications are weighted towards one outcome or another, which is similar to our findings. And the number of physical reclassifications in this paper varies from the approximate number indicated by the model, as is the case for the application by Caudill and Groothuis (2005) to DK responses, which is the only prior example where physical reclassifications are performed. Finally, tests indicate that predicted probabilities from the LCMNL are significantly different than those from the MNL. This appears to contradict previous findings by Caudill, Ayuso, and Guillen (2005) and Caudill and Groothuis (2005) that the estimates from the two models are similar. Because the two models differ significantly, the predicted probabilities from each model are simultaneously tested against those from the data generating process to determine which model is better. The LCMNL appears to outperform the MNL.

By using data created via a known data generating process, this paper has shed new light on the best method for handling ambiguous observations in the empirical analysis. Since the ideal method depends on the source of the ambiguity, future work could assess the technique used by Groothuis and Whithead (2001), where the authors use likelihood ratio tests of multinomial and ordered logit models to determine whether the ambiguous variable represents a distinct or middle response. In addition, the analyses in this paper could be replicated using data generated with inter-variable correlations from different real-world datasets (see equation 11) to see how this influences the results.

References

Abrevaya, Jason, and Jerry A. Hausman. "Semiparametric Estimation with Mismeasured Dependent Variables: An Application to Duration Models for Unemployment Spells." *Annales D'Economie et de Statistique* 55-56 (1999): 243-75.

Alberini, A., K. Boyle, and M. Welsh. "Analysis of Contingent Valuation Data with Multiple Bids and Response Options Allowing Respondents to Express Uncertainty." *Journal of Environmental Economics and Management* 45, no. 1 (2003): 40-62.

Carson, R. T., W. M. Hanemann, R. J. Kopp, J. A. Krosnick, R. C. Mitchell, S. Presser, P. A. Ruud, V. K. Smith, M. Conaway, and K. Martin. "Referendum Design and Contingent Valuation: The Noaa Panel's No-Vote Recommendation (Vol 80, Pg 335, 1998)." *Review of Economics and Statistics* 80, no. 3 (1998): 484-87.

Caudill, S. B. "A Logit Model with Missing Information Illustrated by Testing for Hidden Unemployment in Transition Economies." *Oxford Bulletin of Economics and Statistics* 68, no. 5 (2006): 665 - 77.

Caudill, S. B., M. Ayuso, and M. Guillen. "Fraud Detection Using a Multinomial Logit Model with Missing Information." *Journal of Risk and Insurance* 72, no. 4 (2005): 539-50.

Caudill, S. B., and P. A. Groothuis. "Modeling Hidden Alternatives in Random Utility Models: An Application To "Don't Know" Responses in Contingent Valuation." *Land Economics* 81, no. 3 (2005): 445-54.

Cavanagh, C., and R. P. Sherman. "Rank Estimators for Monotonic Index Models." *Journal of Econometrics* 84, no. 2 (1998): 351-81.

Cramer, J. S., and G. Ridder. "Pooling States in the Multinomial Logit Model." *Journal of Econometrics* 47, no. 2-3 (1991): 267-72.

Dempster, A. P., N. M. Laird, and D. B. Rubin. "Maximum Likelihood Estimation from Incomplete Data Via the Em Algorithm." *Journal of the Royal Statistical Society, Series B* 39 (1977): 1 - 38.

Groothuis, P. A., and J. C. Whitehead. "Does Don't Know Mean No? Analysis of 'Don't Know' Responses in Dichotomous Choice Contingent Valuation Questions." *Applied Economics* 34, no. 15 (2002): 1935-40.

Haener, M. K., and W. L. Adamowicz. "Analysis of 'Don't Know' Responses in Referendum Contingent Valuation Questions." *Agricultural and Resource Economics Review* 27, no. 2 (1998): 219 - 30.

Han, A. K. "Nonparametric Analysis of a Generalized Regression-Model - the Maximum Rank Correlation Estimator." *Journal of Econometrics* 35, no. 2-3 (1987): 303-16.

Hausman, J. A., J. Abrevaya, and F. M. Scott-Morton. "Misclassification of the Dependent Variable in a Discrete-Response Setting." *Journal of Econometrics* 87, no. 2 (1998): 239-69.

Johannesson, M., P. O. Johansson, B. Kristrom, and U. G. Gerdtham. "Willingness to Pay for Antihypertensive Therapy--Further Results." *J Health Econ* 12, no. 1 (1993): 95-108.

Kaiser, H.F., and K. Dickman. "Sample and Population Score Matrices and Sample Correlation Matrices from an Arbitrary Population Correlation Matrix." *Psychometrika* 27 (1962): 179-82.

Powell, J. "Estimation of Semiparametric Models." In *Handbook of Econometrics*, edited by R. Engle and D. McFadden, 2444-521: North Holland, 1994.

Venzon, D. J., and S. H. Moolgavkar. "A Method for Computing Profile-Likelihood-Based Confidence-Intervals." *Applied Statistics-Journal of the Royal Statistical Society Series C* 37, no. 1 (1988): 87-94.

Wang, H. "Treatment of "Don't-Know" Responses in Contingent Valuation Surveys: A Random Valuation Model." *Journal of Environmental Economics and Management* 32, no. 2 (1997): 219-32.

Table 4.1
Summary Statistics

		S=1			S=2			S=3		
Variable	**N**	**Mean**	**Median**	**Std Dev**	**Mean**	**Median**	**Std Dev**	**Mean**	**Median**	**Std Dev**
Y (Measurement Error)	500	1.7480	2.0000	0.7679	1.7700	2.0000	0.7604	1.7580	2.0000	0.7646
Y (Uncertainty)	500	1.7880	2.0000	0.7952	1.6380	2.0000	0.6722	1.5760	1.0000	0.6520
X1	500	0.0684	0.0944	0.9688	-0.1016	-0.1393	1.9823	0.1844	0.3981	2.9421
X2	500	0.0159	-0.0028	0.9935	0.0085	0.0870	2.0359	-0.0688	0.0012	2.9895
X3	500	-0.0259	-0.0796	0.9721	-0.0040	-0.0387	2.0723	-0.1952	-0.1538	3.1397
X4	500	0.0620	0.1155	0.9960	-0.0619	-0.0396	1.9817	-0.0576	0.1496	2.9878
X5	500	0.0273	0.0392	0.9791	-0.0371	-0.0755	1.9431	-0.0961	-0.1457	2.9461
X6	500	0.0003	0.0349	1.0294	0.0083	0.0957	1.9254	-0.2408	-0.2230	2.9406
X7	500	0.0889	0.0594	1.0155	0.0905	0.0269	2.0762	0.0878	0.0315	2.9632
X8	500	0.0588	0.1130	1.0424	0.0539	-0.0097	1.9678	-0.1265	-0.1971	2.8849
X9	500	0.0634	0.1092	0.9536	0.0042	0.0503	1.9834	0.2040	0.2140	2.9842

s=n, implies results are for data created with a population standard deviation of n, where n=1,2,3.

Table 4.2
Reclassification of Ambiguous Observations

Standard Deviation	Method	Outcome 1	2	3	Number Correct
1	$\sum \hat{d}^{*}_{i,3j}$	23.95	42.28	33.77	NA
		(26.68)	(50.61)	(38.71)	(NA)
2		26.87	30.21	42.92	NA
		(4.44)	(19.61)	(30.95)	(NA)
3		32.63	34.90	32.47	NA
		(13.26)	(17.44)	(14.30)	(NA)
1	**Actual**	11	52	37	49/63
		(22)	(49)	(45)	(42/71)
2		24	30	46	51/54
		(2)	(19)	(34)	(15/21)
3		31	35	34	62/66
		(11)	(22)	(12)	(24/33)

Numbers for the ambiguous observations representing respondent uncertainty are in parentheses.

Table 4.3
LCMNL Results (*s*=2)

		Data w/Measurement Error		Data w/Respondent Uncertainty	
Parameter	Y	Estimate	St. Error	Estimate	St. Error
Intercept	2	-0.6841	0.1594	-6.3505	0.2748
Intercept	31	-2.3692	1.0937	-3.2936	0.9193
Intercept	32	-3.0208	0.8889	-7.6228	0.3120
Intercept	33	-1.9246	0.5561	-0.6251	0.3571
X1	2	-0.5229	0.0745	-8.3630	0.0893
X1	33	-0.1829	0.1759	-1.1006	0.2348
X2	2	-0.2206	0.0672	-6.1415	0.0803
X2	33	-0.1318	0.1706	-0.3935	0.1874
X3	2	1.3913	0.0930	22.0237	0.1159
X3	33	0.4741	0.2908	2.3279	0.5329
X4	2	-0.3348	0.0668	-5.6242	0.0773
X4	33	0.0630	0.1487	-0.7604	0.2066
X5	2	0.3278	0.0686	4.9108	0.0778
X5	33	0.1687	0.1482	0.4265	0.2020
X6	2	-0.5531	0.0763	-8.8899	0.0866
X6	33	0.3352	0.2013	-0.4690	0.2186
X7	2	0.3591	0.0642	5.7648	0.0768
X7	33	-0.0160	0.1498	0.5405	0.1917
X8	2	-0.4665	0.0729	-9.4049	0.0893
X8	33	-0.0844	0.2053	-0.8775	0.2462
X9	2	-0.2968	0.0710	-4.5793	0.0845
X9	33	0.1259	0.1620	-0.6252	0.1781

Recall: the parameter estimates for Y=31 and Y=32 are constrained to equal those for Y=1 and Y=2, respectively.

s=2, implies results are for data created with a population standard deviation of 2.

Table 4.4
Traditional Multinomial Logit Results (s=2)

		Data w/Measurement Error				Data w/Respondent Uncertainty		
Parameter	Y	Estimate	St. Error	P Value	Y	Estimate	St. Error	P Value
Intercept	2	-0.4609	0.1589	0.0037	2	-1.1473	0.2793	<.0001
Intercept	3	-0.3945	0.1423	0.0056	3	-0.5699	0.2069	0.0059
X1	2	-0.4446	0.0792	<.0001	2	-1.3091	0.1737	<.0001
X1	3	-0.1954	0.0689	0.0045	3	-0.6969	0.1290	<.0001
X2	2	-0.2074	0.0733	0.0046	2	-0.8071	0.1397	<.0001
X2	3	-0.1471	0.0674	0.0290	3	-0.2103	0.1046	0.0443
X3	2	1.1828	0.1079	<.0001	2	3.3820	0.3389	<.0001
X3	3	0.5860	0.0896	<.0001	3	1.4422	0.2206	<.0001
X4	2	-0.3020	0.0743	<.0001	2	-0.8411	0.1351	<.0001
X4	3	-0.1375	0.0697	0.0484	3	-0.3646	0.1096	0.0009
X5	2	0.2509	0.0769	0.0011	2	0.6893	0.1350	<.0001
X5	3	0.1234	0.0722	0.0873	3	0.1809	0.1116	0.1050
X6	2	-0.4287	0.0811	<.0001	2	-1.3322	0.1756	<.0001
X6	3	-0.0496	0.0717	0.4888	3	-0.5054	0.1276	<.0001
X7	2	0.2926	0.0701	<.0001	2	0.8505	0.1292	<.0001
X7	3	0.0974	0.0652	0.1353	3	0.3109	0.0994	0.0018
X8	2	-0.3822	0.0791	<.0001	2	-1.3450	0.1798	<.0001
X8	3	-0.1557	0.0715	0.0294	3	-0.4781	0.1275	0.0002
X9	2	-0.2400	0.0751	0.0014	2	-0.6545	0.1289	<.0001
X9	3	-0.0496	0.0682	0.4676	3	-0.3684	0.1040	0.0004

s=2, implies results are for data created with a population standard deviation of 2.

Table 4.5

Binary Logit Models (original data)

	s=1			s=2			s=3		
Parameter	Estimate	P Value	Normalized Estimate	Estimate	P Value	Normalized Estimate	Estimate	P Value	Normalized Estimate
Intercept	-0.5351	0.0002	------	-0.7527	0.0002	------	-0.5211	0.0278	
	(0.1442)		------	(0.2016)		------	(0.2368)		
X1	-0.817	<.0001	0.2482	-1.0058	<.0001	-0.3117	-1.1662	<.0001	-0.3758
	(0.1598)			(0.1427)			(0.1723)		
X2	-0.7378	<.0001	0.2241	-0.5607	<.0001	-0.1737	-0.3829	0.0003	-0.1234
	(0.1552)			(0.1132)			(0.1069)		
X3	2.3289	<.0001	0.7074	2.4691	<.0001	0.7651	2.3407	<.0001	0.7543
	(0.2359)			(0.2664)			(0.3119)		
X4	-0.6906	<.0001	0.2098	-0.5742	<.0001	-0.1779	-0.4399	<.0001	-0.1418
	(0.1478)			(0.1090)			(0.0927)		
X5	0.8811	<.0001	0.2676	0.5667	<.0001	0.1756	0.6999	<.0001	0.2255
	(0.1587)			(0.1154)			(0.1122)		
X6	-1.1506	<.0001	0.3495	-0.9382	<.0001	-0.2907	-0.8172	<.0001	-0.2633
	(0.1631)			(0.1412)			(0.1350)		
X7	0.5954	<.0001	0.1808	0.6131	<.0001	0.1900	0.5893	<.0001	0.1899
	(0.1481)			(0.1068)			(0.1079)		
X8	-1.0822	<.0001	0.3287	-0.9394	<.0001	-0.2911	-0.9340	<.0001	-0.3010
	(0.1584)			(0.1411)			(0.1450)		
X9	-0.3169	0.0290	0.0963	-0.4504	<.0001	-0.1396	-0.2725	0.0017	-0.0878
	(0.1451)			(0.1069)			(0.0867)		

Standard errors are in parentheses.

Normalized implies that the sum of the estimates' squares is equal to one.

* implies Ho: βhat = β, rejected at significance level of 10%

** implies Ho: βhat = β, rejected at significance level of 5%

*** implies Ho: βhat = β, rejected at significance level of 1%

s=n, implies results are for data created with a population standard deviation of n, where n=1,2,3.

Table 4.6

Binary Logit Models (without outcome=3 observations)

	s=1				s=2				s=3			
	Data w/Measurement Error		Data w/Respondent Uncertainty		Data w/Measurement Error		Data w/Respondent Uncertainty		Data w/Measurement Error		Data w/Respondent Uncertainty	
Parameter	Estimate	P Value	Estimate	P Value	Estimate	P Value	Estimate	P Value	Estimate	P Value	Estimate	P Value
Intercept	-0.5505	0.0010	-0.7608	0.0010	-0.9420*	<.0001	-1.0750**	<.0001	-0.5644	0.0480	-0.9346	0.0050
	(0.1599)		(0.2302)		(0.2467)		(0.2899)		(0.2848)		(0.3346)	
X1	-0.8555	<.0001	-1.2680	<.0001	-1.0475	<.0001	-1.1497	<.0001	-1.2346	<.0001	-1.4152*	<.0001
	(0.1749)		(0.2739)		(0.1664)		(0.2011)		(0.2016)		(0.2667)	
X2	-0.8041	<.0001	-0.6784	0.0020	-0.5222	<.0001	-0.7521	<.0001	-0.4288	<.0001	-0.6194	<.0001
	(0.1756)		(0.2165)		(0.1312)		(0.1607)		(0.1195)		(0.1653)	
X3	2.2731	<.0001	3.2884***	<.0001	2.4912	<.0001	3.0595**	<.0001	2.4716	<.0001	2.7835	<.0001
	(0.2602)		(0.3987)		(0.3056)		(0.4051)		(0.3708)		(0.4704)	
X4	-0.7169	<.0001	-0.9482**	<.0001	-0.6102	<.0001	-0.7144	<.0001	-0.4930	<.0001	-0.5036	<.0001
	(0.1665)		(0.2264)		(0.1208)		(0.1421)		(0.1090)		(0.1221)	
X5	0.8830	<.0001	0.9098	<.0001	0.5620	<.0001	0.6098	<.0001	0.6855	<.0001	0.7357	<.0001
	(0.1789)		(0.2359)		(0.1258)		(0.1482)		(0.1280)		(0.1455)	
X6	-1.1693	<.0001	-1.5716**	<.0001	-1.0353	<.0001	-1.2447	<.0001	-0.9277	<.0001	-0.9866	<.0001
	(0.1822)		(0.2730)		(0.1723)		(0.2083)		(0.1690)		(0.1871)	
X7	0.5761	0.0010	0.8970*	<.0001	0.6252	<.0001	0.7626*	<.0001	0.6462	<.0001	0.6513	<.0001
	(0.1676)		(0.2283)		(0.1211)		(0.1478)		(0.1272)		(0.1475)	
X8	-1.1121	<.0001	-1.3662*	<.0001	-1.0286	<.0001	-1.1839	<.0001	-1.0387	<.0001	-1.1182	<.0001
	(0.1823)		(0.2426)		(0.1712)		(0.2104)		(0.1756)		(0.2098)	
X9	-0.1535	0.3100	-0.3370	0.1060	-0.4291	0.0010	-0.5071	<.0001	-0.3589	0.0010	-0.3219	0.0080
	(0.1512)		(0.2084)		(0.1302)		(0.1421)		(0.1033)		(0.1216)	

Standard errors are in parentheses.

Normalized implies that the sum of the estimates' squares is equal to one.

* implies Ho: βhat = β, rejected at significance level of 10%

** implies Ho: βhat = β, rejected at significance level of 5%

*** implies Ho: βhat = β, rejected at significance level of 1%

s=n, implies results are for data created with a population standard deviation of n, where n=1,2,3.

Table 4.7

Binary Logit Models (with reclassified data)

	s=1						s=2						s=3					
	Data w/Measurement Error			Data w/Respondent Uncertainty			Data w/Measurement Error			Data w/Respondent Uncertainty			Data w/Measurement Error			Data w/Respondent Uncertainty		
Parameter	Estimate	P Value	Normalized Estimate	Estimate	P Value	Normalized Estimate	Estimate	P Value	Normalized Estimate	Estimate	P Value	Normalized Estimate	Estimate	P Value	Normalized Estimate	Estimate	P Value	Normalized Estimate
Intercept	-0.3206	0.0300	------	-0.5719	0.0070	------	-0.9806**	<.0001	------	-0.9575*	0.0010	------	-0.6674	0.0190		-0.4933	0.0800	
	(0.1477)		------	(0.2126)		------	(0.2458)		------	(0.2836)		------	(0.2841)			(0.2814)		
X1	-0.8937	<.0001	-0.2574	-1.4367*	<.0001	-0.2893	-1.0894	<.0001	-0.3166	-1.1794	<.0001	-0.2809	-1.3083*	<.0001	-0.3676	-1.5229**	<.0001	-0.3826
	(0.1671)			(0.2517)			(0.1661)			(0.1972)			(0.2057)			(0.2541)		
X2	-0.8574**	<.0001	-0.2470	-0.8408	<.0001	-0.1693	-0.5481	<.0001	-0.1593	-0.8024*	<.0001	-0.1911	-0.4507	<.0001	-0.1266	-0.6880	<.0001	-0.1728
	(0.1694)			(0.2136)			(0.1305)			(0.1598)			(0.1215)			(0.1535)		
X3	2.4427	<.0001	0.7036	3.6375***	<.0001	0.7325	2.6066	<.0001	0.7575	3.2328***	<.0001	0.7701	2.6659	<.0001	0.7490	2.9631*	<.0001	0.7443
	(0.2543)			(0.3995)			(0.3105)			(0.4174)			(0.3860)			(0.4543)		
X4	-0.7392	<.0001	-0.2129	-0.9823**	<.0001	-0.1978	-0.6277	<.0001	-0.1824	-0.7514*	<.0001	-0.1790	-0.5065	<.0001	-0.1423	-0.4838	<.0001	-0.1215
	(0.1542)			(0.2024)			(0.1217)			(0.1415)			(0.1097)			(0.1115)		
X5	0.9007	<.0001	0.2594	0.9924	<.0001	0.1998	0.5877	<.0001	0.1708	0.6075	<.0001	0.1447	0.7137	<.0001	0.2005	0.8854	<.0001	0.2224
	(0.1683)			(0.2162)			(0.1260)			(0.1474)			(0.1240)			(0.1551)		
X6	-1.1774	<.0001	-0.3392	-1.8413***	<.0001	-0.3708	-1.0731	<.0001	-0.3118	-1.3227*	<.0001	-0.3151	-1.0007	<.0001	-0.2811	-1.0759	<.0001	-0.2703
	(0.1732)			(0.2632)			(0.1718)			(0.2092)			(0.1749)			(0.1872)		
X7	0.6181	<.0001	0.1781	0.8880*	<.0001	0.1788	0.6348	<.0001	0.1845	0.7956**	<.0001	0.1895	0.7008	<.0001	0.1969	0.6783	<.0001	0.1704
	(0.1541)			(0.2104)			(0.1186)			(0.1479)			(0.1318)			(0.1347)		
X8	-1.1803	<.0001	-0.3400	-1.5441**	<.0001	-0.3109	-1.0454	<.0001	-0.3038	-1.2424	<.0001	-0.2960	-1.1175	<.0001	-0.3140	-1.2333	<.0001	-0.3098
	(0.1728)			(0.2382)			(0.1672)			(0.2122)			(0.1820)			(0.2115)		
X9	-0.1789	0.2210	-0.0515	-0.3816	0.0410	-0.0768	-0.4166	0.0010	-0.1211	-0.5272	<.0001	-0.1256	-0.3723	<.0001	-0.1046	-0.3432	0.0020	-0.0862
	(0.1463)			(0.1871)			(0.1239)			(0.1395)			(0.1032)			(0.1087)		

Standard errors are in parentheses.

Normalized implies that the sum of the estimates' squares is equal to one.

* implies Ho: βhat = β, rejected at significance level of 10%

** implies Ho: βhat = β, rejected at significance level of 5%

*** implies Ho: βhat = β, rejected at significance level of 1%

s=n, implies results are for data created with a population standard deviation of n, where n=1,2,3.

Table 4.8

MRE Models (s=1)

	Original Data		Reclassified Data w/Measurement Error		Reclassified Data w/Respondent Uncertainty	
Parameter	Estimate	95% Confidence Interval	Estimate	95% Confidence Interval	Estimate	95% Confidence Interval
Intercept	------	---------	------	---------	------	---------
	------		------		------	
X1	-0.2373**	[-0.3072, -0.1745]	-0.2645**	[-0.3160, -0.1864]	-0.3066	[-0.3629, -0.2439]
	(0.0397)		(0.0393)		(0.0363)	
X2	-0.2235	[-0.2861, -0.1476]	-0.2415	[-0.3061, -0.1733]	-0.1564	[-0.2172, -0.1214]
	(0.0418)		(0.0405)		(0.0292)	
X3	0.7131	[0.6439, 0.7588]	0.7073	[0.6510, 0.7515]	0.7296	[0.6676, 0.7688]
	(0.0347)		(0.0301)		(0.0301)	
X4	-0.1936	[-0.2672, -0.1458]	-0.2174	[-0.2759, -0.1459]	-0.1895	[-0.2500, -0.1491]
	(0.0371)		(0.0392)		(0.0301)	
X5	0.2596	[0.2012, 0.3331]	0.2362	[0.1935, .3152]	0.1900	[0.1468, 0.2380]
	(0.0402)		(0.0370)		(0.0283)	
X6	-0.3608	[-0.4113, -0.2843]	-0.3369	[-0.4012, -0.2651]	-0.3808	[-0.4340, -0.3122]
	(0.0381)		(0.0409)		(0.0367)	
X7	0.1746	[0.1277, 0.2330]	0.1807	[0.1330, 0.2264]	0.1610	[0.1190, 0.2214]
	(0.0317)		(0.0290)		(0.0302)	
X8	-0.3308	[-0.4003, -0.2689]	-0.3452	[-0.3984, -0.2770]	-0.3181	[-0.3728, -0.2584]
	(0.0398)		(0.0372)		(0.0346)	
X9	-0.1016	[-0.1319, -0.0640]	-0.0550***	[-0.0681, -0.0340]	-0.0710**	[-0.0978, -0.0493]
	(0.0209)		(0.0105)		(0.0151)	

MRE Models (s=2)

	Original Data		Reclassified Data w/Measurement Error		Reclassified Data w/Respondent Uncertainty	
Parameter	Estimate	95% Confidence Interval	Estimate	95% Confidence Interval	Estimate	95% Confidence Interval
Intercept	------	---------	------	---------	------	---------
	------		------		------	
X1	-0.3191	[-0.3574, -0.2655]	-0.3213	[-0.3638, -0.2646]	-0.2823	[-0.3321, -0.2381]
	(0.0280)		(0.0306)		(0.0283)	
X2	-0.1658	[-0.2065, -0.1290]	-0.1434	[-0.1945, -0.1174]	-0.1930	[-0.2241, -0.1488]
	(0.0232)		(0.0230)		(0.0229)	
X3	0.7734**	[0.7336, 0.7971]	0.7634	[0.7259, 0.7920]	0.7661**	[0.7327, 0.8016]
	(0.0195)		(0.0203)		(0.0206)	
X4	-0.1673	[-0.2136, -0.1448]	-0.1672	[-0.2163, -0.1445]	-0.1741	[-0.2178, -0.1451]
	(0.0211)		(0.0222)		(0.0217)	
X5	0.1631**	[0.1336, 0.2123]	0.1685**	[0.1309, .2082]	0.1497***	[0.1095, 0.1790]
	(0.0237)		(0.0236)		(0.0208)	
X6	-0.2882	[-0.3366, -0.2415]	-0.3068	[-0.3642, -0.2511]	-0.3152	[-0.3512, -0.2597]
	(0.0296)		(0.0341)		(0.0274)	
X7	0.1875	[0.1450, 0.2232]	0.1989	[0.1413, 0.2246]	0.1897	[0.1557, 0.2328]
	(0.0228)		(0.0255)		(0.0237)	
X8	-0.2852	[-0.3311, -0.2485]	-0.2988	[-0.3451, -0.2567]	-0.3037	[-0.3407, -0.2440]
	(0.0257)		(0.0266)		(0.0307)	
X9	-0.1355	[-0.1673, -0.1045]	-0.1188	[-0.1533, -0.0856]	-0.1257	[-0.1549, -0.0977]
	(0.0195)		(0.0202)		(0.0179)	

MRE Models (s=3)

	Original Data		Reclassified Data w/Measurement Error		Reclassified Data w/Respondent Uncertainty	
Parameter	Estimate	95% Confidence Interval	Estimate	95% Confidence Interval	Estimate	95% Confidence Interval
Intercept	------	---------	------	---------	------	---------
	------		------		------	
X1	-0.3663**	[-0.4162, -0.3448]	-0.3814**	[-0.4049, -0.3296]	-0.3900***	[-0.4253, -0.3543]
	(0.0224)		(0.0228)		(0.0220)	
X2	-0.1179***	[-0.1560, -0.0936]	-0.1235**	[-0.1615, -0.0926]	-0.1711	[-0.1981, -0.1387]
	(0.0191)		(0.0206)		(0.0178)	
X3	0.7592	[0.7192, 0.7822]	0.7499	[0.7145, 0.7792]	0.7426	[0.7016, 0.7730]
	(0.0189)		(0.0195)		(0.0219)	
X4	-0.1390	[-0.1753, -0.1115]	-0.1413	[-0.1772, -0.1078]	-0.1093***	[-0.1493, -0.0964]
	(0.0193)		(0.0209)		(0.0166)	
X5	0.2204	[0.1839, 0.2704]	0.1958	[0.1559, .2371]	0.2214	[0.1830, 0.2613]
	(0.0256)		(0.0244)		(0.0236)	
X6	-0.2681***	[-0.3016, -0.2143]	-0.2771**	[-0.3213, -0.2295]	-0.2732***	[-0.3007, -0.2198]
	(0.0257)		(0.0272)		(0.0241)	
X7	0.1883	[0.1532, 0.2225]	0.1951	[0.1615, 0.2285]	0.1482	[0.1297, 0.1982]
	(0.0213)		(0.0207)		(0.0212)	
X8	-0.3045	[-0.3310, -0.2579]	-0.3066	[-0.3486, -0.2771]	-0.3179	[-0.3513, -0.2776]
	(0.0220)		(0.0221)		(0.0223)	
X9	-0.0869**	[-0.1056, -0.0672]	-0.0987	[-0.1268, -0.0817]	-0.0922**	[-0.1043, -0.0694]
	(0.0113)		(0.0137)		(0.0108)	

Standard errors are in parentheses.

* implies Ho: βhat = β, rejected at significance level of 10%

** implies Ho: βhat = β, rejected at significance level of 5%

*** implies Ho: βhat = β, rejected at significance level of 1%

s=n, implies results are for data created with a population standard deviation of n, where n=1,2,3.

Table 4.9

Ordered Logit Models (s=1)

	Data w/Measurement Error						Data w/Respondent Uncertainty					
	Original			Reclassified			Original			Reclassified		
Parameter	Estimate	St. Error	P Value	Estimate	St. Error	P Value	Estimate	St. Error	P Value	Estimate	St. Error	P Value
X1	-0.3160***	0.1001	0.0020	-0.4281***	0.1146	<.0001	-0.8641	0.1256	<.0001	-1.3455**	0.1783	<.0001
X2	-0.3768	0.0981	<.0001	-0.4653	0.1115	<.0001	-0.6894	0.1221	<.0001	-0.9847***	0.1655	<.0001
X3	0.9584***	0.1088	<.0001	1.3270***	0.1370	<.0001	2.4729*	0.1910	<.0001	3.7779***	0.3353	<.0001
X4	-0.3877	0.0956	<.0001	-0.4814	0.1124	<.0001	-0.5452	0.1166	<.0001	-0.9506***	0.1619	<.0001
X5	0.4107**	0.0991	<.0001	0.5363	0.1134	<.0001	0.7145	0.1220	<.0001	1.0989**	0.1682	<.0001
X6	-0.4738***	0.0935	<.0001	-0.5119***	0.1057	<.0001	-1.1997*	0.1299	<.0001	-1.8373***	0.2039	<.0001
X7	0.2314***	0.0935	0.0130	0.4185	0.1115	<.0001	0.5583	0.1185	<.0001	0.8397**	0.1582	<.0001
X8	-0.4521***	0.0916	<.0001	-0.5682***	0.1066	<.0001	-0.9482	0.1229	<.0001	-1.4491**	0.1795	<.0001
X9	-0.0504***	0.0999	0.6140	-0.0231***	0.1113	0.8350	-0.3231	0.1167	0.0060	-0.5362	0.1514	<.0001

Ordered Logit Models (s=2)

	Data w/Measurement Error						Data w/Respondent Uncertainty					
	Original			Reclassified			Original			Reclassified		
Parameter	Estimate	St. Error	P Value	Estimate	St. Error	P Value	Estimate	St. Error	P Value	Estimate	St. Error	P Value
X1	-0.2581***	0.0513	<.0001	-0.5685***	0.0778	<.0001	-0.8514	0.0997	<.0001	-1.0483	0.1268	<.0001
X2	-0.1272***	0.0491	0.0100	-0.3717**	0.0707	<.0001	-0.4785	0.0824	<.0001	-0.5766	0.0985	<.0001
X3	0.7455***	0.0640	<.0001	1.5889***	0.1305	<.0001	2.1873	0.1823	<.0001	2.6535**	0.2486	<.0001
X4	-0.1960***	0.0506	<.0001	-0.3867	0.0721	<.0001	-0.5634	0.0861	<.0001	-0.7112**	0.1080	<.0001
X5	0.1486***	0.0522	0.0040	0.3944***	0.0770	<.0001	0.4744**	0.0869	<.0001	0.6031	0.1054	<.0001
X6	-0.2437***	0.0522	<.0001	-0.5575***	0.0803	<.0001	-0.8438	0.1030	<.0001	-1.0095	0.1288	<.0001
X7	0.1842***	0.0478	<.0001	0.4040	0.0681	<.0001	0.5496	0.0793	<.0001	0.6745*	0.0966	<.0001
X8	-0.2153***	0.0519	<.0001	-0.5417***	0.0801	<.0001	-0.8636	0.1027	<.0001	-1.0576	0.1306	<.0001
X9	-0.1515***	0.0496	0.0020	-0.2175	0.0679	0.0010	-0.4674*	0.0827	<.0001	-0.5393**	0.0975	<.0001

Ordered Logit Models (s=3)

	Data w/Measurement Error						Data w/Respondent Uncertainty					
	Original			Reclassified			Original			Reclassified		
Parameter	Estimate	St. Error	P Value	Estimate	St. Error	P Value	Estimate	St. Error	P Value	Estimate	St. Error	P Value
X1	-0.2252***	0.0371	<.0001	-0.4789***	0.0583	<.0001	-1.0368	0.1142	<.0001	-1.3868**	0.1899	<.0001
X2	-0.0760***	0.0331	0.0220	-0.1174***	0.0475	0.0130	-0.4959	0.0820	<.0001	-0.6896	0.1246	<.0001
X3	0.4703***	0.0427	<.0001	0.9597***	0.0863	<.0001	2.1050	0.2070	<.0001	2.6847	0.3340	<.0001
X4	-0.0624***	0.0328	0.0570	-0.1237***	0.0444	0.0050	-0.4501	0.0711	<.0001	-0.5351	0.0967	<.0001
X5	0.1237***	0.0344	<.0001	0.2732***	0.0496	<.0001	0.5824	0.0753	<.0001	0.8016	0.1167	<.0001
X6	-0.1360***	0.0340	<.0001	-0.2781***	0.0501	<.0001	-0.8508	0.1033	<.0001	-0.9954	0.1480	<.0001
X7	0.1711***	0.0354	<.0001	0.3609***	0.0540	<.0001	0.6174	0.0826	<.0001	0.7061*	0.1129	<.0001
X8	-0.2389***	0.0377	<.0001	-0.4499***	0.0573	<.0001	-0.8888	0.1033	<.0001	-1.1350	0.1607	<.0001
X9	-0.0639***	0.0342	0.0620	-0.1092***	0.0469	0.0200	-0.2519	0.0646	<.0001	-0.3540	0.0861	<.0001

* implies Ho: βhat = β, rejected at significance level of 10%

** implies Ho: βhat = β, rejected at significance level of 5%

*** implies Ho: βhat = β, rejected at significance level of 1%

s=n, implies results are for data created with a population standard deviation of n, where n=1,2,3.

Figure 4.1
Contingent Valuation Example of Latent Responses

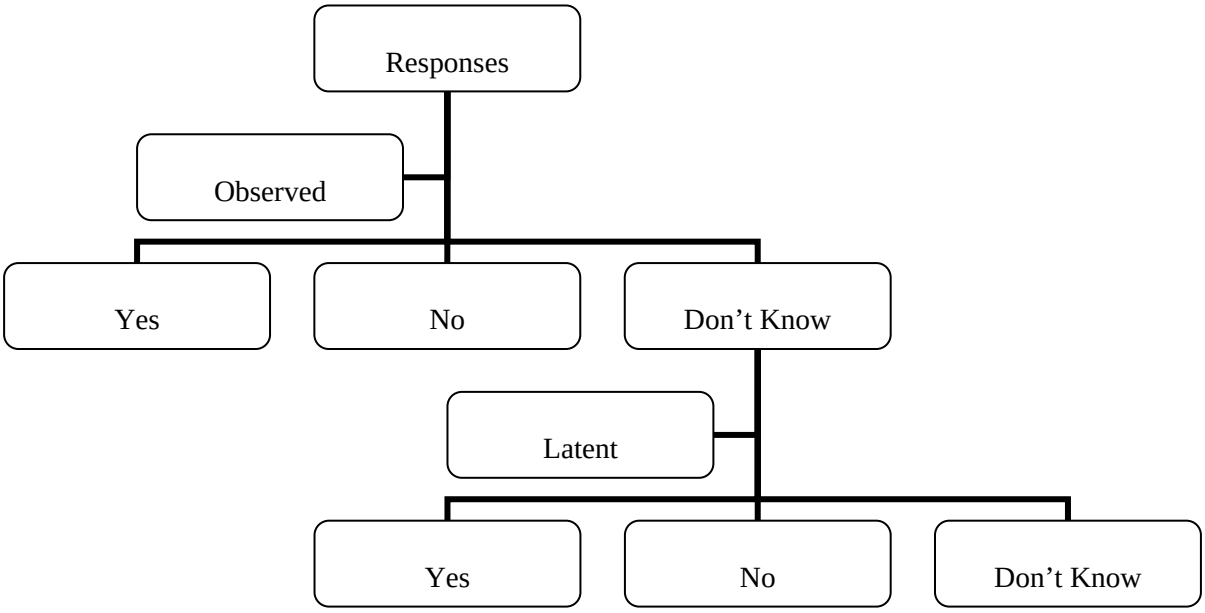


Figure 4.2
Data Generation Process

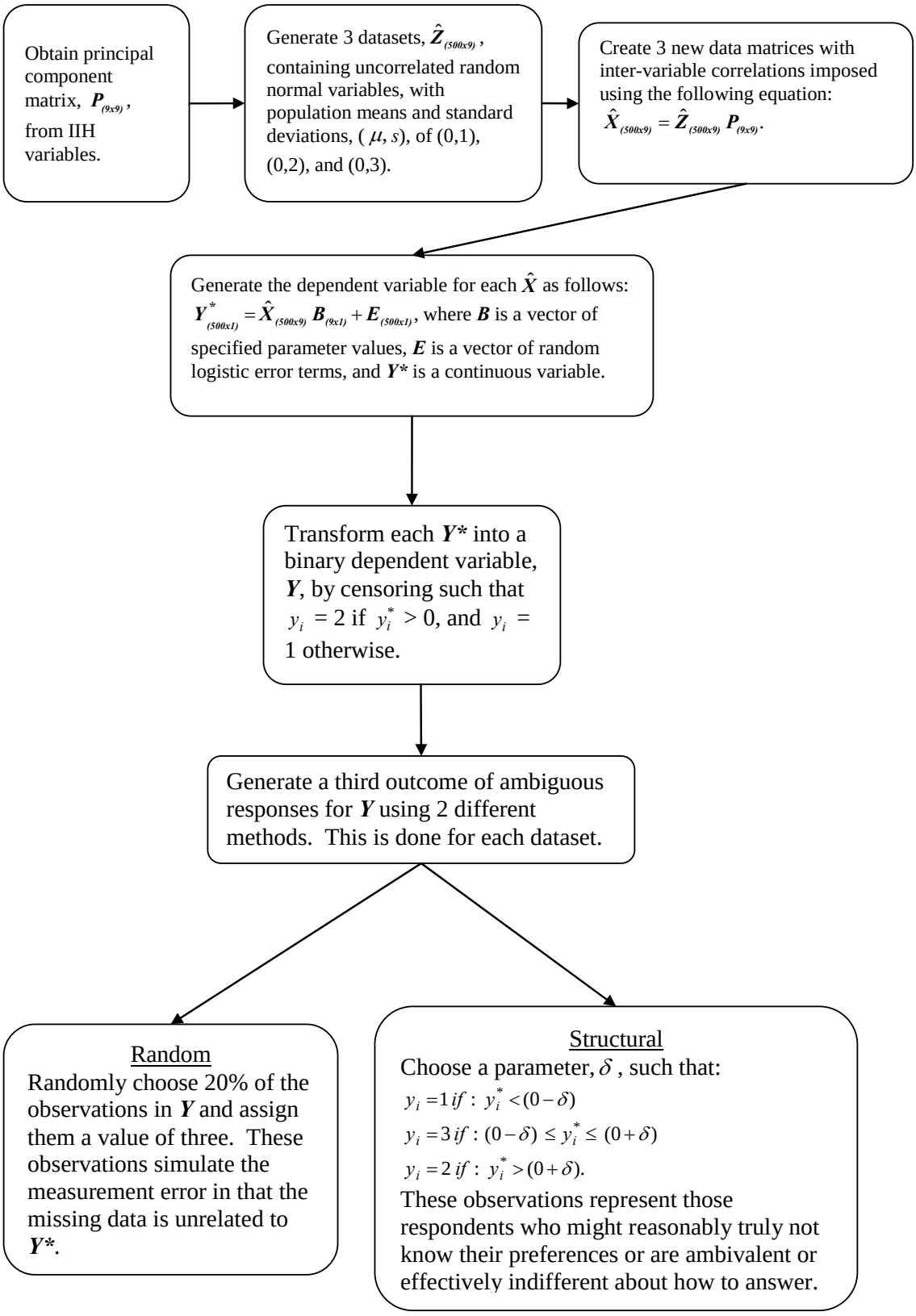
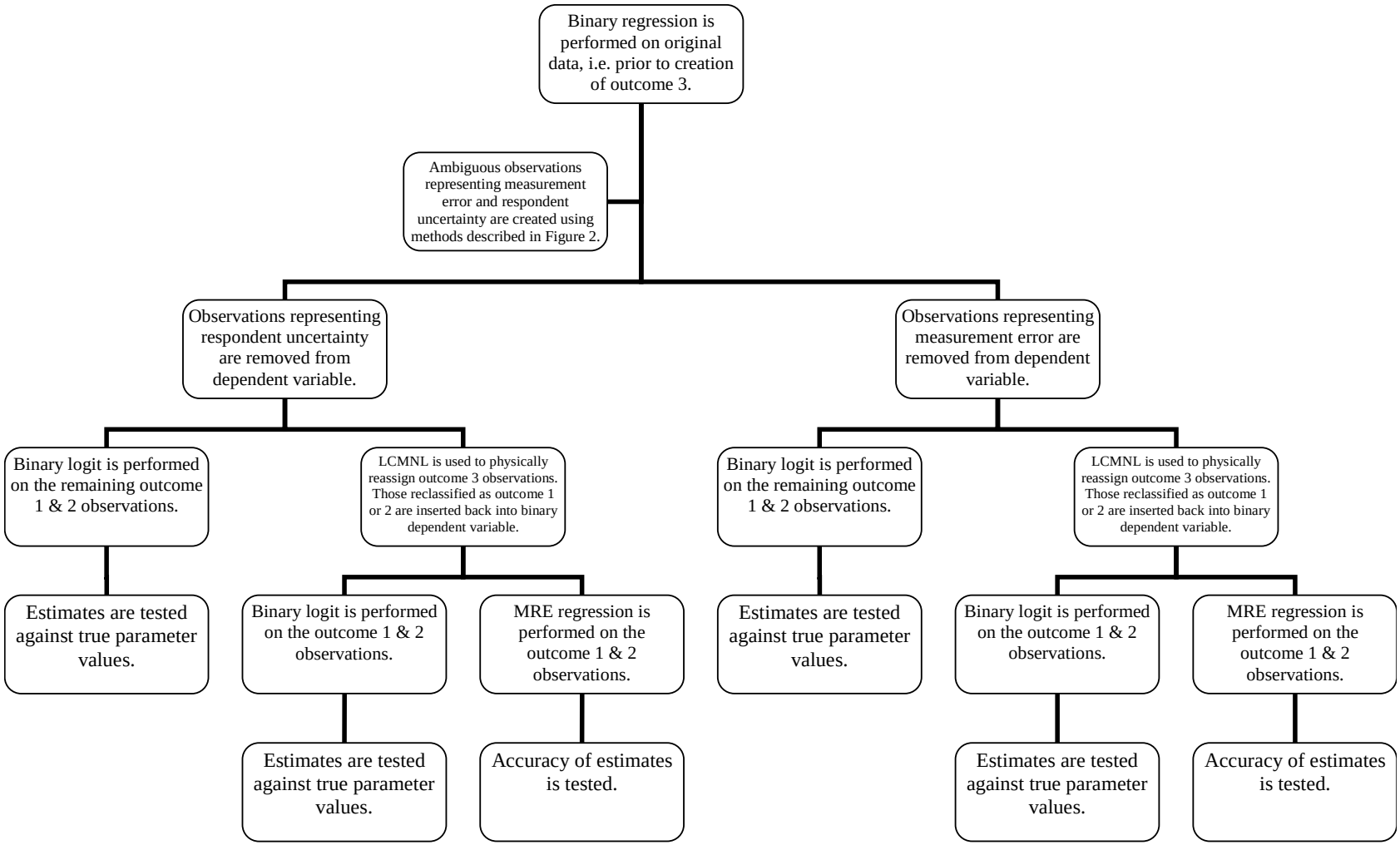


Figure 4.3
Assessment of Ambiguous Observations in Empirical Analysis



Note: This process is repeated for each dataset.

Appendix A

Monotone Rank Estimator (MRE)

The following model for the treatment of mismeasured dependent variables is based on Abrevaya and Hausman (1999), and is an extension of Han's (1987) *generalized regression model.* The true latent dependent variable is represented as follows:

$$Y^* = g(x\beta_0, \varepsilon), \tag{14}$$

where ε is an *i.i.d.* error disturbance, and *g* is an unknown function containing strictly positive partial derivatives at every point. The distribution of *Y* then has the following c.d.f:

$$F_{Y|Y^*}(n|d) = \Pr(Y \le n | Y^* = d), \tag{15}$$

where *n* and *d* represent potential values for the dependent variable. For a model with a binary dependent variable, the probabilities of misclassification are:

$$\phi_0 \equiv \Pr(Y = 1 | Y^* < 0) \tag{16}$$

$$\phi_1 \equiv \Pr(Y = 0 | Y^* > 0)\,. \tag{17}$$

The conditional c.d.f. becomes

$$F_Y(n|Y^*, d < 0) = \begin{Bmatrix} 0 & if & n < 0 \\ 1-\phi_0 & if & n \in [0,1) \\ 1 & if & n \ge 1 \end{Bmatrix} \tag{18}$$

$$F_Y(n|Y^*, d \ge 0) = \begin{Bmatrix} 0 & if & n < 0 \\ \phi_1 & if & n \in [0,1) \\ 1 & if & n \ge 1 \end{Bmatrix}. \tag{19}$$

The MRE, which is a rank estimator for semiparametric monotonic linear index models, is used to estimate the parameters. It consists of a vector $\hat{\boldsymbol{\beta}}^{MRE}$ that maximizes the following objective function:

$$S^{MRE}(b) = \sum_i M(Y_i) \cdot Rank(x_i b) \tag{20}$$

over the set $B \equiv \{\boldsymbol{b} \in \Re^l : |\boldsymbol{b}_l| = 1\}$, where $\Re$ represents the real line, M is an increasing function in Y, $\boldsymbol{X'\beta}$ is the linear index, l represents the number of covariates in x, and $|\boldsymbol{b}_l|$ is the determinant of the $\boldsymbol{b}$ vector. Two comments are in order here. First, note that since the MRE is based on a rank-order process, there is no need to explicitly include an intercept in $\boldsymbol{x_i b}$. Second, equations (18) and (19) imply that the stochastic-dominance conditions are fulfilled when $(1-\phi_0) > \phi_1$, which if it holds implies consistency of the parameter estimates.

The *Rank* function is defined by:

$$x_{i1}b < x_{i2}b < ... < x_{in}b \Rightarrow Rank(x_{im}b) = m. \tag{21}$$

Some examples of functions for M in equation (20) are given by Cavanagh and Sherman (1998). For robustness, $M(Y) = Rank(Y)$, for efficiency $M(Y) = Y$, or an intermediate alternative would be $M(Y) = a\{Y < a\} + Y\{a \le Y \le b\} + b\{Y > b\}$, such that a and b are real numbers and $a < b$. By using a semiparametric approach we may be sacrificing some efficiency relative to a correctly specified parametric model (Powell, 1994); therefore the second option is used to increase the efficiency of the estimates. Finally, the primary condition for consistency is that $E[M(Y)|\boldsymbol{X}]$ is a nonconstant increasing function of

$\boldsymbol{X'\beta}$; however, a sufficient condition for consistency is that the distribution of Y for a higher Y^* first order stochastically dominates that of a Y associated with a lower Y^*.