Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Pallavi S. Mishra-Kalyani Date

Statistical Methods for Causal Inference in Observational Studies

By
Pallavi S. Mishra-Kalyani
Doctor of Philosophy
Biostatistics

Brent A. Johnson, Ph.D.
Advisor

Qi Long, Ph.D.
Advisor

Howard H. Chang, Ph.D.
Committee Member

Jonathan D. Glass, M.D.
Committee Member

Robert H. Lyles, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Statistical Methods for Causal Inference in Observational Studies

By

Pallavi S. Mishra-Kalyani
M.S., Emory University, 2013
S.M., Harvard University, 2009
S.B., Massachusetts Institute of Technology, 2005

Advisors: Brent A. Johnson, Ph.D. and Qi Long, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Biostatistics
2014

Abstract

Statistical Methods for Causal Inference in Observational Studies
By
Pallavi S. Mishra-Kalyani

Observational studies, such as those of patient registries may offer valuable patient and disease information that is impossible to study in a randomized trial, but often pose unique challenges that require special care in estimating a causal effect of treatment. This dissertation is motivated by a registry of patients with amyotrophic lateral sclerosis (ALS) maintained by the Emory ALS Clinic, in which the non-random receipt of the treatment, that is, the insertion of a Percutaneous Endoscopic Gastrostomy (PEG) tube, is time-dependent and both the receipt of treatment and clinical outcomes are subject to "censoring" by death. In order to identify a causal effect of PEG treatment on an outcome, we incorporate and build upon various causal inference methods such as principal stratification and propensity score matching.

After a review of current literature and a more detailed description of the data in Chapter 1, we develop a fully Bayesian modeling approach to estimate the survivor average causal effect (SACE) of PEG on BMI, which is a surrogate outcome measure of nutrition and quality of life, using propensity score methods within a principal stratification framework in Chapter 2. Chapter 3 investigates propensity process matching for estimating treatment effect in observational studies. The Propensity Process is a method that is able to address complex features that are common to observational registries with longitudinally measured data. Matching by Propensity Process outperforms the naive analysis and other non-binary propensity score methods and achieves covariate balance across treatment groups. Chapter 4 extends the methods presented in Chapter 2 to address outcomes that are missing due to a lapse in clinic visits. A single framework incorporating principal stratification using post-treatment survival outcomes, as well as models for the mechanism of missing outcomes and generalized propensity score is used for an unbiased estimation of treatment effect.

Finally, potential future work is explored in Chapter 5. The data of the ALS registry is rich with complications that could inspire new directions of research, and there is significant interest in the issues of observational studies in the statistical community to fuel this methodological research.

Statistical Methods for Causal Inference in Observational Studies

By

Pallavi S. Mishra-Kalyani
M.S., Emory University, 2013
S.M., Harvard University, 2009
S.B., Massachusetts Institute of Technology, 2005

Advisors: Brent A. Johnson, Ph.D. and Qi Long, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Biostatistics
2014

Acknowledgements

I have been fortunate to have had the guidance and support of so many individuals who have been instrumental in the completion of this dissertation.

First and foremost, I want to thank my advisors and committee chairs, Professor Brent Johnson and Professor Qi Long. Each has provided invaluable guidance over the course of my tenure at Emory, and their patience, encouragement, insight, and ideas have greatly enriched my Ph.D. experience and education. They have not only taught me versatile and sophisticated analytical techniques, but also and more importantly have molded me into a scientist with comprehensive research skills. I am honored to call them my advisors and mentors.

I would also like to thank the members of my dissertation committee, Professor Howard Chang, Dr. Jonathan Glass, and Professor Bob Lyles. Their suggestions and helpful discussions have greatly improved my dissertation research and presentation. A special thank you to Dr. Glass and members of the Emory ALS Clinic staff for access to the clinic database and their help providing clinical and biomedical understanding of the disease and patient care.

I am grateful to my colleagues and friends, as well as the department faculty and staff, for making my time at Emory University so rewarding and successful. My time with Department of Biostatistics and Bioinformatics has included some of the most difficult and some of the most joyous moments of my life, and their support through it all has been tremendous.

Lastly, none of this would have been possible without the love and patience of my family and friends. I have been blessed to have five parents throughout my life, Satya, Sushila, Nutan, Dilip and Bindu, and each has taught me and encouraged me as I pursued my goals. All of my siblings, Prashant, Angie, Sachin, and Rita, have cheered me on through every twist and turn and helped

me immensely in the editing process. My brother Prashant, in particular, has inspired me ever since I was a young child by setting the bar for achievement so high. He has never allowed me give up on my dreams no matter how difficult the road ahead seemed. Last but not least, my most heartfelt gratitude goes to my loving and devoted husband, Sameer. He has been my cheerleader, my champion, my best friend, and my partner, and I could not imagine completing this journey without his enthusiasm and support.

Finally, I would like to dedicate this dissertation to my late father, Satya Narayan Mishra. His encouragement was undying and unconditional much like his love, support, and confidence in me. His example taught me strength, patience, and perseverance. I hope one day to fulfill the legacy he left behind, particularly as a parent to my children.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Amyotrophic Lateral Sclerosis and the Motivating Dataset

Amyotrophic Lateral Sclerosis (ALS) is a rare progressive disorder resulting in the degeneration of both upper motor neurons of the cerebral cortex and lower motor neurons of the spinal cord and peripheral nervous system. There are approximately 5000 new diagnoses of ALS in the United States per year, with a total prevalence of about 20,000-30,000 persons in the US and about 6 in 100,000 persons worldwide (of Neurology ALS Work Group 2012, Procaccini & Nemergut 2008, Rowland & Shneider 2001). Males are more frequently affected than females, and 5-10% of all cases are familial, which may be inherited from an autosomal dominant trait. Individual prognosis is difficult to predict, but the range of median survival post onset is approximately 2-3 years, with 20% of patients surviving more than 5 years and 10% of patients surviving more than 10 years post diagnosis (Kiernan et al. 2011, Miller et al. 1999).

While previous research has been inconclusive or inconsistent in determining clinical predictors of survival, older age at diagnosis, female gender, and bulbar onset tend to be negative prognostic factors in analysis of survival in many studies (Gelinas & Miller 2000, Mitchell & Borasio 2007). Additionally, some studies have indicated an association between body mass index (BMI) and survival in ALS patients, with lower BMI associated with shorter survival times. These results indicate that though there is no curative therapy for ALS, maintaining nutrition and weight may be one option for extending survival time or slowing disease progression (Desport et al. 2000, Kasarskis et al. 1996, Muscaritoli et al. 2012, Ngo et al. 2014, Paganoni et al. 2011).

Dysphagia, or difficulty in swallowing, affects almost all patients with ALS, and subsequently along with muscle atrophy and hypermetabolism causes mal-

nutrition amongst this patient population. Nutritional management is key in disease management and palliative care, with enteral nutrition being a long-term option for delivering nutrition (Desport et al. 2006, Muscaritoli et al. 2012). The most common enteral access for medium- and long-term enteral nutrition is percutaneous endoscopic gastrostomy (PEG), which is generally considered when a patient's nutritional status deteriorates and weight loss is greater than 10% of the baseline weight (Goyal & Mozaffar 2014, Park et al. 1992). The American Academy of Neurology has provided practice parameters for the placement of PEG while the forced vital capacity (FVC) of a patient is less than 50% of predicted. This threshold is determined from results of studies showing patients who had diminished vital capacities, FVC <60% of predicted, had shorted survival times than those with greater FVC at the time of procedure (of Neurology ALS Work Group 2012, Goyal & Mozaffar 2014, Gregory et al. 2002).

Many studies show that that PEG can increase food intake and stabilize body weight and BMI effectively as immediate or short-term benefits. However, the results of prolonged survival due to PEG are mixed and many scientist are uncertain of the long term benefits of the procedure (Chio et al. 1999, Kasarskis et al. 1996, Mitsumoto et al. 2003, Shaw et al. 2006, Verschueren et al. 2009). In a recent study, a research group in Italy included a cohort of 150 non-demented dysphagic ALS patients to retrospectively determine the effect of non-randomized PEG insertion on survival using the Kaplan Meier life table method. Though there seemed to be some increase in survival amongst PEG recipients with spinal onset ALS, the authors did not control for selection bias or certain confounders present, causing concern in the interpretation of increased survival (Spataro et al. 2011).

One alternative to PEG is RIG, radiologically inserted gastrostomy, that

can be performed with local anesthesia and fluoroscopic guidance. RIG is associated with lower rates of complications and may be more successful in those individuals with diminished respiratory function. However, RIG also carries a risk of obstruction and dislocation of the tube (Chio et al. 2004).

Data from the Emory ALS Clinic Registry

Our data originates from a registry of patients maintained by the Emory ALS Clinic. The dataset consists of 729 individuals diagnosed with the disease who had one or more visit at the Emory ALS clinic after 1997 and died prior to July 31, 2011. ALS diagnosis was defined as signs of upper and lower motor neuron degeneration in one or more regions in individuals with adult onset symptoms and which could not be attributed to any other disease. Date of death was validated by the Social Security Database.

Table 1.1 describes the overall population characteristics of all 729 individuals in the dataset at initial clinic visit. 520 patients (71%), had a date of diagnosis on the same day as their initial visit to our clinic. Patients visited

Table 1.1: Description of Patient Population at First Clinic Visit (N=729)

	Mean or Proportion	Standard Dev. or N
Has PEG Treatment	43.90%	320
Female	44.99%	328
Spinal Site of Onset	68.04%	496
Diagnosis at First Visit	71.33%	520
FVC at First Visit*	70.72	26.52
NIF at First Visit*	-49.96	15.56
BMI at First Visit*	25.29	5.95
Age at Diagnosis	61.55	12.4
Age at Death	63.33	11.84
Total number of Visits	3.59	2.79

*Not measured for all individuals, N<729

the clinic an average of 3.6 times before death. Mean age is 61.6 years at diagnosis and 63.3 years at death. The difference between age at diagnosis and age at death is consistent with average patient time from diagnosis to death in the general population. Less than half of the patients received PEG at sometime before death (43%), with average time from diagnosis to treatment of 288 days. Most patients have spinal onset of disease (68%), but the minority of patients are female (45%). Though not all patients have recorded body mass index (BMI), forced vital capacity (FVC), or negative inspiratory force (NIF) at baseline, the means from non-missing measurements for these clinical markers are 25.3 kg/m^2 (normal range 18.5 - 24.9 kg/m^2 depending on age), 70.7 percent predicted (normal range >80 percent predicted), and -50.0 $cm\ H_2O$ (normal range >-60 $cm\ H_2O$).

The treatment of surgical insertion of a percutaneous endogastrostomy (PEG) tube was offered to individuals to supplement or be a primary source of nutrition. Though conclusive corroborated evidence of a significant positive benefit of PEG on quality of life or survival outcome has not been identified, many neurologists have noted anecdotal evidence from their medical practice that patients with PEG fare better than those without citepmiller1999. One possibility is that scientific research has been overshadowed by the many complications arising in this patient population when attempting to identify a true treatment effect for PEG.

Kaplan-Meier curves comparing survival time of patients receiving and not receiving PEG in our data are available in Figure 1.1. In each of these plots, patients are considered treated if the PEG tube is inserted prior to the corresponding time point. Though there does not seem to be any survival advantage from PEG treatment in this analysis, this may be because it is unadjusted for potential confounding covariates or selection bias.

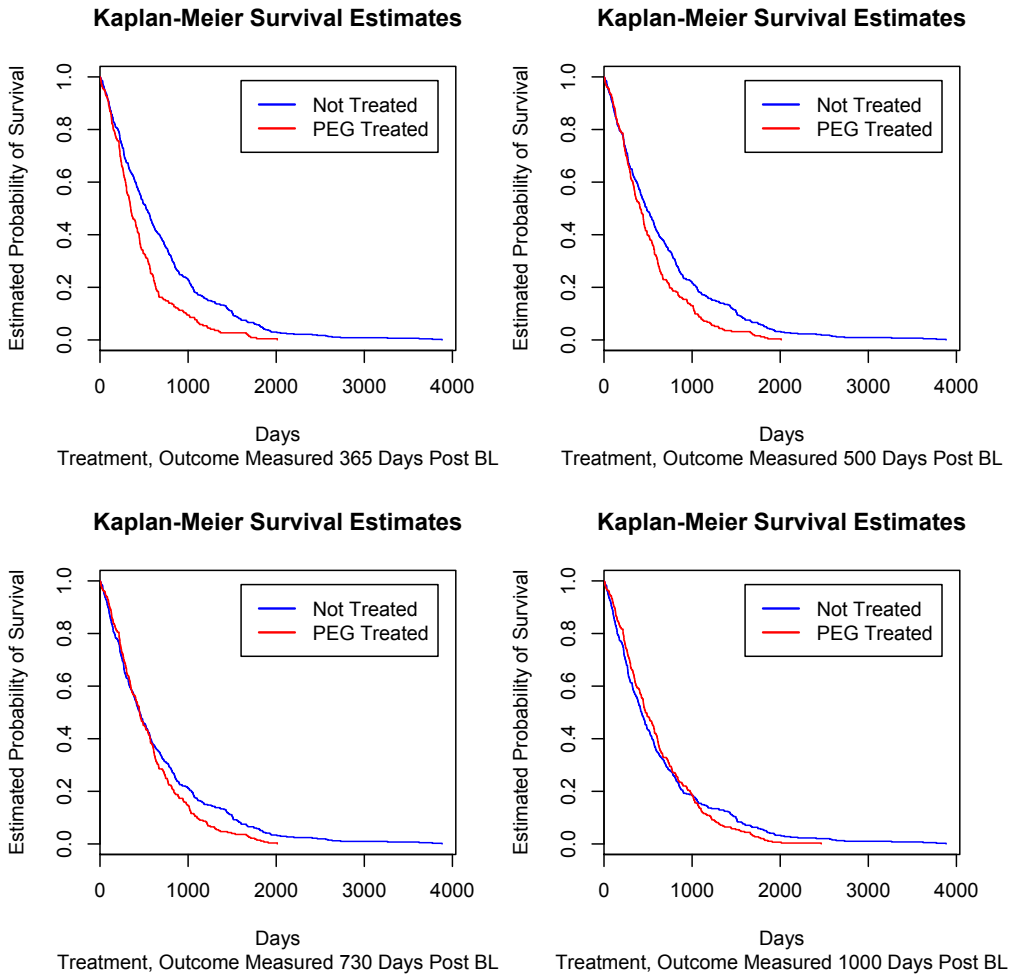

Figure 1.1: Kaplan Meier Curves for Treated and Untreated Populations at Several Time-points of Follow-up (N=729)

1.2 Literature Review

1.2.1 Rubin's Causal Model

The inherent biases that exist in non-randomized or poorly randomized sample require methods from the field of Causal Inference for unbiased estimation of treatment effect. Rubin's Causal Model (RCM) is one important and widely used framework for identifying a causal effect of treatment by means

of potential outcomes. The original methodology was developed by Rubin in 1974, then defended and given its current name by Holland in 1986 (Rubin 1974, Holland 1986).

Potential Outcomes and Assumptions

Potential outcomes are defined as the outcomes Y_i for each unit i with and without treatment, $Y_i(1)$ and $Y_i(0)$. One of these potential outcomes is observed, $Y_i(Z_i)$, and the other is considered a counterfactual outcome. As a result, we are unable to determine the treatment effect on a single unit i, $Y_i(1)$ - $Y_i(0)$. Instead, we find the average treatment effect or average causal effect $ACE = E(Y_i(1) - Y_i(0))$ by making some assumptions about how the treatment is assigned and how it may effect the outcome.

Rubin's model assumes that there is an assignment mechanism that determines the treatment assignment of each individual, thereby determining which of the potential outcomes are observed. In randomized studies, the assignment mechanism can generally be ignored as the design of the study dictates that treatment is unrelated to individual characteristics; however in observational studies, it is assumed that the assignment mechanism is non-random. Thus, it is essential to control for the assignment mechanism when attempting to identify a causal effect. In order to do this, we must rely on certain assumptions.

The first assumption that must be employed for RCM is the Stable Unit Treatment Value Assumption (SUTVA), which states that there is no interference between observed treatment Z_i^{obs} of one unit i and the potential outcomes of $S_{i\prime}$ and $Y_{i\prime}$. Additionally, SUTVA allows for the treatment of all units to be comparable by assuming that there is no variation in the treatment. The assumptions in addition to SUTVA that may be made to estimate the average

treatment effect are listed below.

1. Ignorability of treatment assignment: $Y(1), Y(0) \perp Z$. This assumption may be made conditionally on covariates, the control for treatment assignment by propensity score methods $Y(1), Y(0) \perp Z|X$.
2. Exclusion restriction. Treatment assignment is unrelated to the potential outcomes, once observed treatment is taken into account.
3. Non-zero causal effect of treatment.
4. Monotonicity, or that for unit i $Y_i(1) > Y_i(0)$. This also indicates that when principal stratification is present (described in forthcoming sections), defiers do not exist.

Estimation of Causal Effect

With these assumptions and the general rules of probabilities, the average treatment effect simplifies to $E(Y_i(1)) - E(Y_i(0))$. When random treatment assignment can be assumed, or more commonly in observational studies ignorability of treatment assignment can be assumed by means of conditioning on covariates, the average outcome in the treated or control groups are the average outcome for the whole population if treated or untreated respectively. Thus, the average causal effect would be $ACE = E_1(Y_i(1)) - E_0(Y_i(0))$ or $ACE = E_1(Y_i(1)|X) - E_0(Y_i(0)|X)$ respectively.

There are a number of methods used to estimate the average causal effect, including a modeling, inverse probability weighting, and use of propensity scores. These methods may be used in tandem or individually to remove bias and estimate causal effects. While the basic assumptions for each of the methods are similar (e.g. no unmeasured confounders, correctly specified models), each method acts in unique ways to estimate the unbiased ACE.

A parametric model can be used to identify a causal effect while controlling for confounding variables by means of covariate inclusion, stratification, or weighting. Direct modeling of the average causal effect of treatment can be considered for continuous outcome models by means of maximum likelihood or Bayesian inference for a linear model. Control for confounding covariates is essential, and can be accomplished by techniques such as matching or inclusion as linear predictors. Models for potential outcomes may also be considered, in which case the missing values of counterfactual outcomes may be imputed directly or methods such as principal stratification may be employed.

Inverse probability weighting (IPW) uses the probability of observed treatment given confounding covariates to weight observations. Treated individuals have a weight of the conditional probability of treatment, $P(Z = 1|X_i)$, and untreated individuals are weighted by the probability of not being treated, $P(Z = 0|X_i)$. After weighting, the treated and untreated populations are considered exchangeable, and thus the weighted averages of the means of each group can be used to calculate the ACE. Because the probability of observed treatment must be modeled to determine the weight, IPW works well with dichotomous or categorical treatment assignments, and can be extended to situations of time to event and repeated measures data. However, this method poses difficulties when considering continuous treatment assignments (Robins et al. 1994, Hernán & Robins 2006, Cole & Hernán 2008).

Propensity scores are a common method for controlling for issues such as confounding and selection bias when estimating a causal effect. Propensity scores offer a single dimension to be conditioned upon for exchangeability of treatment groups, which can be a major advantage when many confounders are present. These scores can be calculated and utilized in many ways, as described in detail in the next section.

1.2.2 Propensity Scores

Propensity scores are balancing scores that are most often used to control for various types of bias in observational studies, including but not limited to selection bias and confounding. They also can be used to test for ignorable treatment assignment, an important assumption of Rubin's Causal Model. The first methods for propensity scores were developed by Rosenbaum and Rubin (1983), but the generation of propensity scores and their application to various types of data have been a key area of interest for Causal Inference research (Rosenbaum & Rubin 1983).

The general idea of any balancing score is conditionally remove any inherent differences between groups. The balancing score $b(X)$ is a function of the observed covariates X such that X is independent of treatment Z conditional on $b(X)$. Rosenbaum and Rubin present five theorems to support the use of propensity scores and other balancing scores, summarized below.

1. The propensity score $e(X)$ is a balancing score.
2. Any score finer than the propensity score, such as $b(X) = X$, is also a balancing score.
3. If treatment assignment is strongly ignorable given X, then it is also strongly ignorable given a balancing score $b(X)$.
4. At any value of a balancing score, comparison of means of an outcome in treated and untreated groups is an average treatment effect, if strongly ignorable treatment is met. This also indicates that use of balancing scores for matching, subclassification, and covariate adjustment produces unbiased treatment effect estimates, so long as treatment assignment is strongly ignorable.

5. Sample estimates of balancing scores produces sample balance on X.

While there are many advantages to using propensity scores, the major disadvantage to these methods is that bias is that one must assume that there are no unmeasured confounders of the treatment effect. In other words, propensity score methods can only account for confounding by covariates that are observed. This assumption is quite strong, especially in data that is poor in covariate measurements.

Types of Propensity Scores

Most of the early research using propensity scores focused on binary treatments. In these cases, the definition of the propensity score is the probability of treatment conditional on X, $b(X) = P(Z = 1|X)$. This probability is generally found using a logistic regression model for treatment using patient covariates as described in equation 1.1. In addition to being used in the manners that Rosenbaum and Rubin described (matching, subclassification or stratification, and covariate adjustment), the traditional propensity scores may also be used for inverse probability of treatment weighting in the likelihood for the outcome variable Y.

$$P\left(Z = 1|X\right) = \frac{e^{X\beta}}{1 + e^{X\beta}} \tag{1.1}$$

However, when the treatment is non-binary, the probability of treatment may either be difficult to quantify or is not the best balancing score. Zhao, Imai, and Van Dyk (2012) have compared two options for flexible propensity scores that have been proposed in earlier literature. The two proposed solutions to non-binary treatment propensity scores are the generalized propensity score (GPS) of Hirano and Imbens (2004) and the P-Function of Imai and Van Dyk (2004). GPS is defined as $e_\psi(X) = p_\psi(Z = z|X)$, and is equal to the treat-

ment assignment model evaluated for observed treatment and covariates. The P-Function uses the entire conditional probability density and requires the existence of a uniquely parameterized propensity function such that $e_\psi(|X)$ depends on X only through $\theta_{psi}(X)$. One example of $\theta_\psi(X)$ is the linear predictor $X_i\beta$ (Hirano & Imbens 2004, Imai & Van Dyk 2004, Zhao et al. 2012).

While both the GPS and the P-Function are valid methods for propensity score methods for non-binary treatments, each has its' own set of strengths and weaknesses. Both require many strong parametric assumptions. The GPS allows for the direct estimation of a dose response function[1], but is less robust in it's modeling than the P-Function. P-Functions are theoretically advantageous, as it achieves independence of outcome Y and treatment Z in low dimensional data, but the model can be prone to mis-specification. Also, it is noteworthy that when covariates are time-independent, the GPS and P-function are equivalent. That is, when these methods are used to calculate a propensity score for covariates measured single time point, such as baseline, the values of the propensity score are equivalent.

Use of Propensity Scores

Perhaps the most common use of propensity scores is for matching treated individuals to untreated individuals for comparison. The advantages and disadvantages of matching mimic those of generally using propensity scores in controlling for bias, with the additional advantage that matching assumes no strictly linear relationships between the outcome and propensity score (Rosenbaum & Rubin 1985*b*).

Several matching algorithms have been discussed in the literature, with

[1]Zhao, Imai, and Van Dyk suggest that the P-Function included in a smooth coefficient model may also be capable of estimating a dose response function

no consensus to a frontrunner among the methods. Examples of matching structures that algorithms include nearest neighbor matching using some distance metric to find the treated and untreated units with a minimal distance between them and stratification matching within some subgroups defined by propensity score. While more complex structures are available, nearest neighbor matching is most often employed for its ease of use and interpretation.

An extension of using propensity scores for matching is stratification or subclassification by propensity scores. In this method, the population are broken down into k strata, which are identified by similar propensity score values. Methods for estimating a causal effect of treatment on outcome are performed within each stratum, and then an overall estimate of effect can be calculated as a weighted average of the within strata effects.

Each of these proposed propensity scores can also be included in the model of the response variable. The standard propensity score can be included as a model covariate, and can also be transformed or manipulated with basis splines to allow flexibility in the relationship between the outcome variable and the propensity score. GPS is included as a linear term, a quadratic term, and an interaction term with treatment (with treatment also having a quadratic term, with a resulting treatment effect outcome estimated as a dose response function (DRF). P-Functions have more flexibility in their utilization, in that they can be included directly as a linear predictor but can also be stratification of the model and or as a smooth coefficient model. The effect of the treatment on outcome when using the P-Function is the average causal effect (ACE).

Finally, propensity scores can be used for weighted likelihood or regression analysis. The most common weighting is known as Inverse Probability weighting (IPW) or Inverse Probability of Treatment weighting (IPTW)(Keisuke Hirano 2003, Robins 1999). The general idea behind this technique is that by

weighting the likelihood by in inverse probability of treatment, the sample is standardized for an unbiased estimate of effect for the overall sample representative population. Implementation of weighting has been discussed widely with implications for model misspecification of the treatment assignment model and the use of time dependent treatment assignment mechanisms (Robins 1999, Rotnitzky & Robins 1995).

1.2.3 Principal Stratification

Often when comparing treatment effects on outcome, other surrogate outcomes and post-treatment characteristics S_i^{obs} are measured and available for analysis. Frangakis and Rubin (2002) suggest that the estimating the effect of treatment within a group of individuals who have the same measure of S_i^{obs} will result in a causal estimand. They formally define principal stratification in two steps. First, the basic principal stratification P^0 is the partition of units i = 1,2,...,n, such that within any P^0 all units will have the same vector $S_i(0), S_i(1)$. $S_i(0)$ and $S_i(1)$ are potential outcomes of S_i that vary by treatment assignment, where only one value is actually observed ($S_i(Z_i)$). Second, principal stratification P with respect to post treatment variables S is a partition of the units whose sets are unions of sets in the basic principal stratification P^0 (Frangakis & Rubin 2002*a*).

Noteworthy properties of principal stratification include: (1) treatment assignment does not affect the stratum S_i^P, (2) an "exclusion" assumption can be made so that if treatment assignment does not effect S_i it must not effect the final outcome, and (3) any principal effect (as defined above) is a causal effect. The first property, which states that the individual strata assignment is unaffected by treatment, is very similar to and can be considered an extension

of SUTVA, which states that there is no interference between observed treatment Z_i^{obs} and the potential outcomes. The second property allows for the principals of instrumental variables to be utilized when estimating the causal effect within a particular stratum of interest (Angrist et al. 1996, Rubin 2005).

The principal effect would then be the comparison of outcome under various treatment alternatives within a principal stratum $S_i^P = S$, specifically the effect estimates between the sets $\left[Y_i\left(1\right) : S_i^P\right]$ and $\left[Y_i\left(0\right) : S_i^P\right]$. However, this treatment effect may not be estimable in every strata. Principal effect of treatment on outcome cannot be measured if individuals in the strata, drops out of a study, or does not comply to treatment. Thus the principal or causal effect of treatment is only estimated in the stratum S_i^P where $S_i(1) = 1$ and $S_i(0) = 1$ (Frangakis & Rubin 2002*a*).

Types of Stratification

Principal stratification has been used to address issues in multiple areas of research by utilizing applicable post-treatment characteristics. Perhaps the most notable area that utilizes principal stratification is effect estimation when non-compliance is present. In fact these methods were used to deal with issues of compliance prior to Frangakis and Rubin's paper formally described principal stratification for various post-treatment characteristics (Imbens & Rubin 1997, Little & Rubin 2000, Yau & Little 2001). Since non-compliance can effectively invalidate randomization of treatment assignment, it is important to appropriately control for various levels of compliance. Principal stratification allows for the identification of four levels of compliance: always takers, compliers, defiers, and never takers. These four groups are defined whether or not an individual abides by the assigned treatment. This method of stratifica-

tion allows for the identification of the treatment effect among compliers, also known as the Complier Average Causal Effect (CACE), and is employed by many when dealing with study data affected by non-compliance (Little et al. 2009, Long et al. 2010).

Principal stratification can also be used to categorize patients by other post-treatment characteristics or surrogate outcomes. For example, Zhang et al stratify by employment status to estimate the effect of a job training program on wages. In this case, the causal effect of the program on wages can only be identified in the group that is employed with or without treatment (Zhang et al. 2009). Their methods are considered to be an extension of principal stratification methods for censoring by death, which generally involve stratification using survival status.

Survivor Average Causal Effect

When individual survival is the characteristic upon which principal stratification is framed, the principal effect in the for the stratum where all outcomes can be observed is termed the Survivor Average Causal Effect (SACE)(Egleston et al. 2009, Hayden et al. 2005, Rubin 2005). In this scenario, the four principal strata are described with regards to survival status and outcome measurement in the list below. Outcome measurement can only be measured on the real set of numbers $\mathbb{R}$ if a patient is alive, and thus those who are not alive would have an outcome measurement that we assume exists in some extended sample space $\{\mathbb{R}, *\}$. It is important to note that time of study follow-up for outcome and survival measurement time must be finite, and so principal strata may only be defined at any given measured time t^* before the end of the follow-up period.

- $LL = \{i|S_i(1) = 1, S_i(0) = 1\}$, or those individuals who would be alive at time t^* regardless of treatment. These individuals would have outcome measurements $Y_i(1), Y_i(0) \in \mathbb{R}$.
- $LD = \{i|S_i(1) = 1, S_i(0) = 0\}$, or those individuals who at time t^* would be alive if they receive the PEG tube but would not be alive if they did not. These individuals would have outcome measurements $Y_i(1) \in \mathbb{R}, Y_i(0) = *$.
- $DL = \{i|S_i(1) = 0, S_i(0) = 1\}$, or those individuals who at time t^* would not be alive if they receive PEG treatment, but would be alive if they did not. These individuals would have outcome measurements $Y_i(1) = *, Y_i(0) \in \mathbb{R}$.
- $DD = \{i|S_i(1) = 0, S_i(0) = 0\}$, or those individuals who would not be alive at time t^* regardless of treatment. These individuals would have outcome measurements $Y_i(1), Y_i(0) = *$.

The survival status in each bullet above are potential outcomes of survival, as opposed to the observed survival status of each individual, which would definitely depend on their observed treatment. Because these potential outcomes are unknown, we use observed treatment and survival at time t^o for prediction of the principal strata in the population. We may accomplish this by determining the mixture of strata that may be present in each group of observed treatment and survival, and then using this information to use standard mixture model analysis for the outcome to estimate the survivor average causal effect, $SACE = E\left(Y_{LL}(1) - Y_{LL}(0)\right)$. The SACE is the principal effect measured in the LL stratum, as this is the only stratum in which a well-defined effect of treatment on outcome is available.

Monotonicity

Because principal stratification is a technique used within Rubin's Causal Model for estimation of unbiased effects, the assumptions outlined in Section 1.2.1 are valid and employed. Of these assumptions, monotonicity is of particular interest in the principal stratification framework. Monotonicity assumes that there are no "defiers." For example, if principal stratification is based on patient survival (S_i^{obs} is death or censoring by death), then defiers would be those individuals who die if treated but live if not treated: $S_i(1) = 0$ and $S_i(0) = 1$. This assumption need not be applied for every study that includes principal stratification, as the data may have characteristics that allow for defiers. Continuing the above example of principal stratification based on survival outcomes, consider study in which the treatment is very dangerous or strenuous on the patient, such as an invasive surgery. It is possible that a subpopulation of weak patients would be not be able to withstand treatment or recovery and may die shortly after treatment, but could stay alive it left untreated. For such a data analysis, it may be appropriate to allow all four principal strata to exist and to relax the assumption of monotonicity.

Missing Data

Missing outcomes can be a major complication in the estimation of a causal effect. A simple solution may be to drop all records with a missing outcome, but there are draw backs to this approach. One issue is that removing these individuals from a analysis assumes that the missing outcomes are missing completely at random, which is a very strong assumption about the missing mechanism. Also, when dropping all individuals with missing outcomes from the analysis, a considerable amount of information contained in these records

may be lost in the process.

One option in the realm of causal inference methodology is to make assumptions about missingness within the principal stratification framework. For example, Frumento et al. (2012) utilize an assumption originally made by Frangakis & Rubin (1999) in a compliance based principal stratification framework. This assumption, termed latent ignorability, states that the missing mechanism is ignorable conditional on principal strata membership. If no other assumptions are made, no model is needed for the missing mechanism once the latent principal strata are known. However, Frumento et al. also make assumptions of exclusion restrictions, which limit the probabilities of missingness conditional on values of treatment and post-treatment variables. Given these exclusion restrictions, it is necessary to include the probability of missingness in the observed data likelihood. This probability may be modeled using known parametric regression forms and with covariate data if necessary. The resulting parameter estimates and probabilities are not generally of causal interest, but are rather are regarded as nuisance parameters.

1.2.4 Data Augmentation Algorithm

In Bayesian Modeling, one generally wants to sample from a posterior density. However, in real world applications, missing or unobserved data often prevents direct sampling from this posterior. Tanner and Wong (1987) developed the Data Augmentation (DA) algorithm, a technique similar to the EM algorithm in that we impute missing data or unobserved information from observed data to estimate parameters from a posterior distribution.

Assume data Y is observed and augmented by the latent variable Z. Y has a distribution depending on θ, and so our objective is to sample from the

posterior density $P(\theta|Y)$. However, we may only be able to sample from the augmented posterior $P(\theta|Y,Z)$ and the predictive distribution $P(Z|Y,\theta)$. The DA algorithm provides and iterative process for sampling from these two estimable distributions to approximate the posterior for θ.

The iterative process of the DA algorithm has two steps. The first step is the Imputation or I-Step, in which using the current estimate of $\theta^{(k)}$, Z is generated from $P(Z|Y,\theta^{(k)})$. The imputed Z is then used to update the approximation of $P(\theta|Y)$ in the Posterior or P-step. These two steps are repeated until there is convergence of the posterior distribution approximation(Tanner & Wong 1987).

Chapter 2

Estimating treatment effect in observational studies in the presence of censoring due to death

2.1 Introduction

Data from clinics and disease registries can offer an opportunity to examine the effect of treatment over time in a natural setting. However, the nature of such data can pose many challenges for unbiased estimation of a treatment effect, particularly in the case of diseases with high mortality rates such that "censoring by death" is a concern. In this paper, we are interested in estimating a causal effect of treatment in a clinical registry of patients with amyotrophic lateral sclerosis (ALS), a neurodegenerative disorder with a very poor prognosis (Gelinas & Miller 2000, Procaccini & Nemergut 2008). However, the data feature several unique and noteworthy characteristics that require special consideration. The most obvious difficulty is that non-random treatment assignment may lead to issues of selection bias and confounding. Secondly, the data are collected longitudinally for each individual, but with varying time elapsed between each individual's measurements and amongst all individuals. Finally, the fatal and fast-progressing nature of ALS results in the potential censoring by death of the outcome; in other words, outcome is only observed if a patient survives beyond the date on which outcome in measured.

The data in the ALS registry is measured longitudinally at patient clinic visits occurring at uneven intervals of time for each individual, $i = 1, 2, ..., N$. The baseline visit is defined as the first clinic visit for all patients, and is denoted by t_0. The set of patient characteristics, denoted by $\mathbf{D}$, includes both a set of characteristics measured only at baseline and a set that are measured at each clinic visit. Of particular interest in this data is the effect of a surgical insertion of a percutaneous endoscopic gastrostomy (PEG), a palliative procedure that provides enteral nutrition, on the outcome BMI, a proxy measure of adiposity associated with nutritional status and mortality. Denoted as $\mathbf{Y}$, the observed outcome is the measurement of BMI collected at the clinic visit that

is closest to but not past 1 year post-baseline (t^o). Time of survival, denoted by T_S, is measured as days from baseline until death and is used to create an indicator of survival S within the first year post-initial clinic visit. For those individuals who receive PEG prior to 1 year post-baseline, treatment is recorded as both the time from PEG surgery until 1 year post baseline, $\mathbf{T_Z}$, as well as a binary indicator of PEG surgery within the first year post baseline, $\mathbf{Z}$.

The use of principal stratification in the context of Rubin's Causal Model while considering unobserved outcome data as censored by death provides a framework for analysis of the ALS registry data. The principal stratification framework is described in detail by Frangakis & Rubin (2002*a*), and Zhang & Rubin (2003) extend this methodology for stratification when the outcome is "censored" by a post-treatment variable such as survival or graduation. Zhang et al. (2009) further outlines specific parametric approaches for identification of survivor average causal effect (SACE) in the analysis of truncation by death using principal stratification. The unbiased estimation of a principal effect, defined as the causal effect within principal strata, relies on an assumption of ignorablility or no unmeasured confounding (Frangakis & Rubin 2002*a*). However, when selection bias or confounding may be present, either as residual confounding in a randomized clinical trial or due to observational data, Schwartz et al. (2012) show that the resulting principal effect estimate is likely to be biased. This result indicates that in the absence of randomization or when the randomization scheme results in poor balance among treatment groups, there is a need to incorporate methods for alleviating selection bias and confounding within a principal stratification framework.

There are many methods that address selection bias or confounding in observational studies, including the propensity score introduced by Rosenbaum

& Rubin (1983), which provides a means of balancing covariates across treatment groups thereby mimicking a randomized study design. Though propensity scores were introduced for balancing treatment assignment groups when treatment is binary, other authors have extended these methods to non-binary treatment assignment models, such as generalized propensity score methods (Imai & Van Dyk 2004, Hirano & Imbens 2004). Specifically, these methods allow for the estimation of causal effects when treatment assignment is ordinal, categorical, continuous, semi-continuous, or even multivariate. In the proposed methodology, the generalized propensity score is estimated using a proportional-hazards model for time to treatment.

In the following sections, we propose a framework for the estimation of a causal effect of treatment that combines principal stratification with adjustment of covariates by means of generalized propensity score. Although Jo & Stuart (2009) included propensity scores within a principal stratification framework, the scores were utilized not for removal of selection bias or confounding in the randomized data of interest, but rather for prediction of principal strata membership for a matched analysis. Furthermore, propensity score methods for non-binary treatment assignment have not yet been employed for conditional ignorability in a principal stratification framework. The methodology presented in this paper addresses the practical application of the principal stratification framework to observational data. The proposed methodology is described in detail in Section 2.2, followed by an application to the data from the Emory ALS Clinic registry in Section 2.3. Simulation studies are presented in Section 2.4.

2.2 Methodology

For data such as the Emory ALS Clinic registry, where treatment assignment can be measured as both a binary indicator of treatment and as time of treatment, different models may be postulated for the effect of treatment on outcome varying by the treatment variables used. In particular, in the proposed methodology, a dichotomous treatment model and a time of treatment model are considered. The dichotomous treatment model considers the binary definition of PEG treatment only, while the time of treatment model incorporates both the dichotomous definition of treatment as well as a measure of time elapsed from treatment until the time of outcome measurement ($t^o - T_Z$). In each of these models, the definition of treatment assignment not only effects the interpretation of the estimated treatment effect but also the modeling and calculation of the propensity score.

Patient characteristics in combination with time to treatment, indicator of no treatment, and the propensity score vector **PS**, as described in the following section, comprise the matrix of observed data **X**, which is used in parts for modeling. In the stratified regression model for the outcome **Y**, the subset of **X** that is included in the analysis is $\mathbf{X}_1$, which may include $\mathbf{T_Z}$, **Z**, **PS**, and a subset of variables from **D**, depending on the outcome model and the stratum. $\mathbf{X}_2$ is the subset of **X** included in the regression model for the principal strata probabilities $P(G = g)$, which may include **PS** and a subset of variables from **D**.

2.2.1 Framework for Causal Inference

Two commonly used assumptions are made in this causal inference framework. First, we make the Stable Unit Treatment Value Assumption (SUTVA)

as defined by Cox in 1958 and summarized by Rubin (1980). This assumption states that there is no interference among the potential outcomes of one individual and the treatment choices of another individual. Secondly, we assume Strong Ignorability of Treatment Assignment (Rosenbaum & Rubin 1983), which states the distribution of the potential outcomes is independent of treatment assignment, given the observed covariates. The ignorability assumption often proves to be non-trivial, particularly for data from observational studies.

As earlier defined, the indicator for treatment from baseline until t^o (time of outcome measurement) is $\mathbf{Z}$, the outcome of interest is $\mathbf{Y}$, and the post-treatment variable of survival is $\mathbf{S}$. Using the Rubin Causal Model (Holland 1986) as a framework for causal inference, we can define potential outcomes $\mathcal{Y} = \{Y_i(z^P), z^P \in \mathcal{Z} \text{ for } i = 1...n\}$ and $\mathcal{S} = \{S_i(z^P), z^P \in \mathcal{Z} \text{ for } i = 1...n\}$, where $\mathcal{Z}$ is the set of potential treatment values and $Y_i(z^P)$ and $S_i(z^P)$ are the potential outcomes for a given potential treatment z^P.

2.2.2 Generalized Propensity Scores

When considering only the effect of a dichotomous treatment, standard propensity scores may be employed. The individual propensity scores are estimated using a logistic regression model for the probability of treatment using all available baseline patient characteristics, $\mathbf{D}$. Specifically, we model $logit\left(P\left(Z=1\right)\right) = \mathbf{D}^T\boldsymbol{\beta}$, and use the parameter estimate $\hat{\boldsymbol{\beta}}$ to calculate the individual probabilities $P\left(Z_i = 1\right)$ as each individual's propensity score, PS_i.

However, when considering a model for the outcome that identifies associations with time of treatment, we must consider a more flexible model for propensity score. The generalized propensity score methods proposed by Imai & Van Dyk (2004) and Hirano & Imbens (2004) allow the inclusion of the

information provided by covariates and more importantly control for selection bias and confounding when a non-binary treatment assignment model is considered in a non-randomized sample (Zhao et al. 2012). Additionally, Imai and Van Dyk derive "large-sample" theoretical results of balancing properties and ignorable treatment assignment that resemble the results for the standard propensity score proposed by Rosenbaum & Rubin (1983). Therefore, for time to treatment, we may use a Cox proportional hazards model $h(t) = h_0(t) \exp\left(\mathbf{D}^T \boldsymbol{\beta}\right)$, and the estimated linear predictor $\mathbf{D}_i^T \hat{\boldsymbol{\beta}}$ is the generalized propensity score.

The estimated propensity scores, as defined for dichotomous treatment model and for the time of treatment model, are included in both the model for the outcome ($\mathbf{Y}$) as well as the model determining principal strata ($\mathbf{G}$), to control for issues of selection bias and confounding. It is noteworthy that inclusion of the propensity scores in the principal strata model is necessary in the absence of randomization, as otherwise the principal effect is likely to be biased. To allow flexibility in the control for selection bias when including each of the propensity scores a linear predictor, the use of quadratic and cubic polynomial higher order terms are also considered for each propensity score.

2.2.3 Principal Stratification

If a patient is not alive under the treatment that is actually received, $S_i = 0$, the outcome Y_i cannot be measured. We may consider those outcomes that are not measured due to patient death as not defined on the set of real positive numbers, $\mathbf{R}^+$. Following the notation of Zhang & Rubin (2003), we can instead consider the non-observed outcomes to be *, extending our sample space to $\{\mathbf{R}^+, *\}$. In the presence of this censoring of the outcome by death, principal

stratification using post treatment survival status allows for estimation of the treatment effect. Specifically, the Survivor Average Causal Effect (SACE) is defined as the mean difference in the outcomes of treated individuals compared to untreated individuals in the LL stratum, $E(Y_{LL,i}(1)) - E(Y_{LL,i}(0))$. The four potential principal strata are constructed by pairing indicators of survival by treatment scenarios at time of outcome measurement, as defined below.

- $LL = \{i|S_i(1) = 1,\ S_i(0) = 1\}$, or those patients who would be alive at time t^o regardless of treatment.
- $LD = \{i|S_i(1) = 1,\ S_i(0) = 0\}$, or those patients who at time t^o would be alive if they receive the PEG tube but would not be alive if they did not.
- $DL = \{i|S_i(1) = 0,\ S_i(0) = 1\}$, or those patients who at time t^o would not be alive if they receive PEG treatment, but would be alive if they did not.
- $DD = \{i|S_i(1) = 0,\ S_i(0) = 0\}$, or those patients who would not be alive at time t^o regardless of treatment.

The probabilities of the four strata (π_{LL}, π_{LD}, π_{DL}, and π_{DD}) can be modeled with a multinomial logit model using a subset of the observed covariate matrix $\mathbf{X}$, $\mathbf{X_2}$, which must include $\mathbf{PS}$ and may include patient characteristics $\mathbf{D}$. The probability of an individual being in principal strata g is given in equation (2.1). As in any multinomial logit model, one category must be selected as a reference group.

$$\pi_{g,i} = P(G_i = g) = \frac{\exp\left(X_{2i}^T \boldsymbol{\alpha}_g\right)}{\sum_{g'} \exp\ \left(X_{2i}^T \boldsymbol{\alpha}_{g'}\right)} \tag{2.1}$$

However, for any given individual i, we only observe the survival outcome given the observed treatment status. These four groups based on observed

data are defined as follows:

- $O(1,1) = \{i|Z_i = 1,\ S_i^{obs} = 1\}$: individuals who are treated and are alive at time t^o
- $O(1,0) = \{i|Z_i = 1,\ S_i^{obs} = 0\}$: individuals who are treated and are not alive at time t^o
- $O(0,1) = \{i|Z_i = 0,\ S_i^{obs} = 1\}$: individuals who are not treated and are alive at time t^o
- $O(0,0) = \{i|Z_i = 0,\ S_i^{obs} = 0\}$: individuals who are not treated and are not alive at time t^o

These observed groups are composed of mixtures of the principal strata. In other words, $O(1,1)$ is comprised of a mixture of individuals from the LL and LD strata, $O(1,0)$ is comprised of individuals from the LL and DL strata, $O(0,1)$ is comprised of individuals from the DD and DL strata, and $O(0,0)$ is comprised of individuals from the DL and DD strata.

Thus far, the framework for principal stratification does not employ an assumption of monotonicity. This assumption implies that the DL stratum (those who do not live with the receipt of treatment, but will live if untreated) does not exist. Other than the reduction of principal strata to three rather than four, the model framework is largely the same under the monotonicity assumption. For completeness of methodology and to test the sensitivity of the results to this assumption, results for all data analysis and simulation studies are reported with and without the monotonicity assumption.

2.2.4 Bayesian Framework for Estimation and Inference

The observed outcome Y_i, which is only observed when $S_i = 1$, is assumed to have a normal distribution, f_g, within each of the principal strata and with

Table 2.1: Individual observed likelihood by observed treatment and survival group

			G_i			
Observed Group	Z_i	S_i	LL	LD	DL	DD
O(1,1)	1	1	$\pi_{LL,i} f_{LL,i}$	$\pi_{LD,i} f_{LD,i}$	-	-
O(1,0)	1	0	-	-	$\pi_{DL,i} f_{DL,i}$	$\pi_{DD,i} f_{DD,i}$
O(0,1)	0	1	$\pi_{LL,i} f_{LL,i}$	-	$\pi_{DL,i} f_{DL,i}$	-
O(0,0)	0	0	-	$\pi_{LD,i} f_{LD,i}$	-	$\pi_{DD,i} f_{DD,i}$

parameters and covariates that differ by strata. Specifically, the outcome distributions are defined as $Y_{g,i} \sim N(\mathbf{X}_{1,g}\boldsymbol{\eta}_g, \sigma_g^2)$ for $g \in LL, LD, DL$. $\mathbf{X}_{1,LL}$ includes the column for intercept, one or both of the treatment variables depending on the treatment assignment model considered, and the estimated propensity score corresponding ot the treatment assignment model. The outcome models for the LD and DL strata do not include any treatment covariates as the individuals with an observed outcome in each of these strata are either all treated or all untreated, respectively. Therefore, $X_{1,LD}$ and $X_{1,DL}$ include columns for intercept and propensity score only.

Using the stratified distributions and the probability of each principal stratum, the structure of the observed data likelihood for any individual and for all possible combinations of Z_i and S_i is given in Table 2.1. Each cell value is the likelihood of the observed data if the values of the individuals' strata are known. Thus the conditional probability of $G_i = g$ given the observed data is the ratio of each cell to the total of that row. Rows $O(1,0)$ and $O(0,0)$ are included in this table for a comprehensive understanding of the possible combinations of treatment and survival, but individuals who fall into these groups do not have outcome data that will contribute to the observed data likelihood since $S_i = 0$ and thus Y_i is unobserved. Therefore, individuals in this group will only contribute to the model for the probability of principal strata, which is reflected in the observed data likelihood in the Appendix I.

Prior distributions for the specified parameters in the observed data likelihood should be chosen carefully, with thought to distributions that may be informative, proper, and conjugate where appropriate. For this analysis, conjugate multivariate normal and inverse-gamma distributions are assigned for the prior distributions of the different forms of $\boldsymbol{\eta}_g$ and σ_g^2 respectively. The prior distributions for $\boldsymbol{\alpha}_g$ are non-informative and are proportional to 1. Details of each prior distribution are provided in Appendix II.

$$
\begin{aligned}
&P(\theta|Y,S,Z,G,D,PS) \propto P(\theta)P(Y|S,Z,G,D,PS) \\
&\propto \sigma_{LL}^{2(-\nu_{LL}-1)} exp\left(-\frac{\omega_{LL}}{\sigma_{LL}^2}\right) |\sigma_{LL}^2 V_{LL}|^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_{LL}^2}(\eta_{LL}-\mu_{LL})^T V_{LL}^{-1}(\eta_{LL}-\mu_{LL})} \\
&\times \sigma_{LD}^{2(-\nu_{LD}-1)} exp\left(-\frac{\omega_{LD}}{\sigma_{LD}^2}\right) |\sigma_{LD}^2 V_{LD}|^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_{LD}^2}(\eta_{LD}-\mu_{LD})^T V_{LD}^{-1}(\eta_{LD}-\mu_{LD})} \\
&\times \sigma_{DL}^{2(-\nu_{DL}-1)} exp\left(-\frac{\omega_{DL}}{\sigma_{DL}^2}\right) |\sigma_{DL}^2 V_{DL}|^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_{DL}^2}(\eta_{DL}-\mu_{DL})^T V_{DL}^{-1}(\eta_{DL}-\mu_{DL})} \\
&\times \prod_i \Bigg\{ I_{(G_i=LL)} \left(\frac{e^{X_{2i}\alpha_{LL}} \sigma_{LL}^{-1} e^{\frac{\left(Y_i - X_{1,LL,i}\eta_{LL}\right)^2}{2\sigma_{LL}^2}}}{1 + e^{X_{2i}\alpha_{LL}} + e^{X_{2i}\alpha_{DL}} + e^{X_{2i}\alpha_{DD}}} \right) \\
&\times I_{(G_i=LD)} \left(\frac{\sigma_{LD}^{-1} e^{\frac{\left(Y_i - X_{1,LD,i}\eta_{LD}\right)^2}{2\sigma_{LD}^2}}}{1 + e^{X_{2i}\alpha_{LL}} + e^{X_{2i}\alpha_{DL}} + e^{X_{2i}\alpha_{DD}}} \right) \\
&\times I_{(G_i=DL)} \left(\frac{e^{X_{2i}\alpha_{DL}} \sigma_{DL}^{-1} e^{\frac{\left(Y_i - X_{1,DL,i}\eta_{DL}\right)^2}{2\sigma_{DL}^2}}}{1 + e^{X_{2i}\alpha_{LL}} + e^{X_{2i}\alpha_{DL}} + e^{X_{2i}\alpha_{DD}}} \right) \\
&\times I_{(G_i=DD)} \left(\frac{e^{X_{2i}\alpha_{DD}}}{1 + e^{X_{2i}\alpha_{LL}} + e^{X_{2i}\alpha_{DL}} + e^{X_{2i}\alpha_{DD}}} \right) \Bigg\}
\end{aligned}
\tag{2.2}
$$

The posterior distribution of the parameters given the observed data likelihood and specified prior distributions is provided in equation (2.2). Though the principal stratum of each individual is unknown, the observed treatment and survival groups may be used to inform imputation of the principal strata assignments. One option in Bayesian analysis is the Data Augmentation (DA)

algorithm (Tanner & Wong 1987), which treats G as missing data, imputes G, and subsequently simulates the posterior distributions of $\boldsymbol{\theta}$, a given imputed G.

The DA algorithm is employed by using two iterative and alternating steps to simulate a complete data likelihood and allow for posterior inference. The first step, the Imputation or I-step, imputes the value of the principal strata G_i for each individual. This is accomplished by using the parameter values $\boldsymbol{\alpha}_g^{(k)}$, $\boldsymbol{\eta}_g^{(k)}$, and $\sigma_g^{2(k)}$ from the current approximation of posterior (from the kth iteration) to generate $G_i^{(k+1)}$ by using the conditional probabilities that are given by taking the ratio of cell value to row total in Table 2.1. The conditional probabilities, $\rho_{O,i}$ are used in a Bernoulli distribution that imputes individual membership to one of the two principal strata that correspond with the observed group O (see Appendix III). More specifically, at the (k+1) iteration, each individual has a probability of being in a stratum that depends on their observed values (Z_i, S_i, Y_i, PS_i).

The P-step, or Posterior step, is then employed by using the imputed complete data set, and the parameters $\boldsymbol{\theta}^{(k)} = \left(\pi_g^{(k)}, \boldsymbol{\eta}_g^{(k)}, \sigma_g^{2(k)}\right)$ can be updated to $\boldsymbol{\theta}^{(k+1)} = \left(\pi_g^{(k+1)}, \boldsymbol{\eta}_g^{(k+1)}, \sigma_g^{2(k+1)}\right)$ by sampling from the full conditional distributions of each parameter, or the distribution in which all other parameters in $\boldsymbol{\theta}$ are conditioned upon. Either the Gibbs Sampler or the Metropolis-Hastings (MH) Algorithm may be employed for sampling, with choice of algorithm influenced by the type of full conditional distribution. The full conditional distributions of each parameter (given the imputed G at each iteration k) are provided in Appendix A.4.

2.3 Application to Emory ALS Clinic Data

The ALS registry dataset includes data from 729 patients who visited the Emory ALS clinic at least once from January 1, 1997 and July 31, 2011. All individuals were diagnosed with ALS prior to first clinic visit, and none had received PEG treatment. Of these patients, 38 patients were excluded for not having any follow up clinic visits within 1 year of their first clinic visit, 25 were excluded for having extremely long survival times (>5 years post-baseline), and 86 were excluded for having no post-baseline measurements of the outcome BMI within the first year of follow up. Characteristics measured at baseline for each individual include sex, site of ALS onset, age at diagnosis, BMI at baseline, and days from diagnosis to first clinic visit (ΔT_{DX}). Additionally, some individual characteristics are measured at each clinic visit including forced vital capacity (FVC), change in FVC from baseline (ΔFVC_{t^o}), change in BMI from baseline (ΔBMI_{t^o}), and total number of clinic visits. Those characteristics that are measured as continuous variables (namely age at diagnosis, BMI at baseline, FVC, and change in FVC from baseline) are normalized before inclusion as covariates for the propensity score model, principal strata model, or outcome model.

A comparison of those who receive treatment within one year of follow-up and those who do not among the remaining 580 individuals in the ALS registry is available in Table 2.2. Of the 200 treated patients 41.5% or 83 individuals are alive one year from baseline, while of the 384 untreated individuals 54.2% or 206 individuals are alive at this time-point ($p < 0.01$). In general, treated individuals tend to have characteristics that align with greater risk of advanced disease, such as smaller values of mean FVC, increased age, lower proportions of spinal onset, and higher proportions of females.

Baseline measurements of BMI are not significantly different among the

treated and untreated populations were not significant at a level of $\alpha = 0.05$, however, FVC measurements taken at baseline are significantly different ($p < 0.01$), with treated individuals having a lower mean measurement than untreated individuals. Demographic characteristics such as age at diagnosis and sex are also significantly different among treated and untreated, with treated individuals being about 3 years older and more likely to be female than untreated individuals ($p < 0.01$ and $p = 0.02$ respectively). Additionally, the proportion of patients with spinal onset of disease is significantly lower in the treated population ($p < 0.01$), which along with the lower mean FVC at baseline, higher age, and greater proportion of females indicates increased risk of advanced disease in the treated patient population.

When considering time of outcome measurement, 1 year post-baseline, the clinical measurements of FVC in patients who are treated and untreated have an even greater gap than the measurements at baseline ($p < 0.01$). This result may imply that those individuals who do receive treatment within one year post-baseline are not only in poorer condition at baseline, but they are generally in poorer condition as time and their disease progresses. BMI is also lower for patients who are treated when compared to untreated ($p = 0.05$).

Comparison of treatment groups are available at additional points of time elapsed post-baseline in Table A.1 of the Appendix. Though some of the differences in means or proportions are not significant at other time points, the general trend in characteristics in each population is consistent. Treated individuals tend to have characteristics that align with greater risk of advanced disease, such as smaller values of mean FVC, increased age, lower proportions of spinal onset, and higher proportions of females.

2.3.1 Balance of Covariates

To test the balance of covariates using propensity score methods, the association of patient characteristics with the indicator of treatment was examined for change after conditioning on the proposed propensity scores. Table 2.3 presents the p-values of these associations when examined marginally, conditioned on the standard propensity score as calculated from a logistic regression of the treatment indicator, and conditioned on the generalized propensity score as calculated from a Cox proportional hazards regression of the time to treatment.

Overall, the use of propensity scores does balance the covariates across treatment groups. Most patient characteristics have a significant marginal association with treatment. After conditioning on the standard propensity score, all associations of patient characteristics with treatment become non-significant, indicating that balance of covariates is successfully achieved. Conditioning on the generalized propensity score is also fairly successful in balancing covariates, with most patient characteristic associations with treatment

Table 2.2: Comparison of PEG treated and untreated populations 1 year post-baseline (N=580)

	Treated (N=200)		***Untreated (N=380)***		
	Mean/P	*SD or n*	*Mean/P*	*SD or n*	*p-value*
BMI at t^o	23.87	5.52	24.89	5.63	0.05
Baseline (BL) BMI	24.92	5.89	25.62	5.68	0.19
ΔBMI_{t^o},	-1.13	2.12	-0.65	2.92	0.04
FVC at t^o	46.84	21.44	63.49	25.11	< 0.01
Baseline (BL) FVC	65.24	25.58	73.98	25.84	< 0.01
ΔBMI_{t^o}	-18.86	22.09	-10.42	16.96	< 0.01
Age at Diagnosis	64.91	10.15	61.91	12.17	< 0.01
$\Delta T_{DX} > 30$ days	0.21	41	0.19	71	0.68
Proportion Surviving	0.42	83	0.54	206	< 0.01
Prop. of Females	0.52	103	0.41	156	0.02
Prop. of Spinal Onset	0.43	86	0.80	304	< 0.01

becoming non-significant. Though two patient characteristics (baseline FVC and number of visits in the first year of follow-up post-baseline) remain associated with treatment after conditioning on the generalized propensity score.

2.3.2 Estimation of SACE of PEG Treatment

SACE of PEG Treatment is estimated in 16 model scenarios; in addition to the two definitions of treatment (binary indicator of treatment and time of treatment), models are considered for various levels of propensity score inclusion (none, linear, quadratic, and cubic propensity score terms) and with or without the assumption of monotonicity. For all analyses, the MCMC algorithm was run for a total of 10,000 iterations, with a burn-in period of 5000 iterations.

Table 2.4 presents a comparison of models with the linear propensity score and without propensity score terms. Overall, the inclusion of a propensity score may change the magnitude, direction and, significance of the treatment effect estimates. In the case of the binary treatment indicator only model, though the treatment effect remains negative and non-significant, the magnitude of the effect of treatment on BMI is closer to zero after including propensity scores. For comparison, Table 2.5 includes the results of the

Table 2.3: Balance of Covariates Among Treatment Groups: P-values of Treatment Indicator Effect on Patient Characteristics after Inclusion of Propensity Scores (N=491)

	Without PS	Standard PS	Generalized PS
Baseline BMI	0.186	0.499	0.185
Baseline FVC	<0.001	0.807	0.002
Age at Diagnosis	0.003	0.457	0.278
Number of Visits	<0.001	0.188	0.007
Spinal Site of Onset	<0.001	0.587	0.113
Female Sex	0.016	0.131	0.717

PEG treatment effect estimates in all eight variations of the binary treatment model. When all four strata are considered, the treatment effect is similar when any propensity score terms are included in the model. The application of the monotonicity assumption does cause some minor discrepancies in the magnitude and direction of the effect estimates in the models with propensity score terms. However, all treatment effect estimates are non-significant as all credible intervals include the null value of 0. Therefore, there seems to be no significant treatment effect in the model with binary indicator of PEG treatment only.

The time of treatment model, which includes both time from treatment to 1 year post-baseline and a binary indicator of treatment, shows a change not only in the magnitude but also in the significance of the effect estimates of PEG treatment when comparing the results from inclusion of linear propensity score term to that without a propensity score term (Table 2.4). Most notably, when propensity scores are used, the effect of the PEG treatment indicator is larger in magnitude and is significant, compared to a smaller non-significant estimate when no propensity score terms are included in the model. These results remain significant after including higher order propensity score terms and employing the monotonicity assumption (Table 2.6).

Table 2.4: SACE of PEG treatment (with 95% credible intervals)on BMI measured 1 year post-baseline (N=491)

	No Propensity Score		Linear PS Model	
	Mean	*95% CI*	*Mean*	*95% CI*
Outcome model including binary treatment indicator only				
Peg Treatment	-1.21	(-2.51, 0.07)	-0.17	(-1.56, 1.22)
Linear PS Term	-	-	-4.20	(-6.72, -1.70)
Outcome model including both binary indicator and time of treatment				
Time from Treatment to t*	-0.49	(-0.77, -0.21)	-0.34	(-0.60, -0.06)
Peg Treatment	1.84	(-0.18, 3.85)	2.69	(0.82, 4.53)
Linear PS Term	-	-	-1.30	(-1.91, -0.67)

Table 2.5: Comparison of SACE estimates of PEG treatment (with 95% credible intervals) with and without the monotonicity assumption in the dichotomous treatment model (N=491)

	PEG Treatment Effect Estimate	
	Monotonicity	*All Four Strata*
No Propensity Score	-1.37 *(-2.69, -0.02)*	-1.21 *(-2.51, 0.07)*
Linear Propensity Score Term	0.40 *(-0.93, 1.73)*	-0.17 *(-1.56, 1.22)*
Quadratic Propensity Score Terms	0.37 *(-0.99, 1.70)*	-0.15 *(-1.57, 1.27)*
Cubic Propensity Score Terms	-0.24 *(-1.66, 1.20)*	-0.16 *(-1.59, 1.25)*

Table 2.6: Comparison of SACE estimates of PEG treatment (with 95% credible intervals) with and without the monotonicity assumption in the time of treatment model (N=491)

	PEG Treatment Indicator		Time from Treatment to t*	
	Monotonicity	*All Strata*	*Monotonicity*	*All Strata*
No Propensity Score	1.48 *(-0.71, 3.57)*	1.84 *(-0.18, 3.85)*	-0.50 *(-0.19, -0.80)*	-0.49 *(-0.21, -0.77)*
Linear PS Term	2.25 *(0.19, 4.30)*	2.69 *(0.82, 4.53)*	-0.35 *(-0.05, -0.65)*	-0.34 *(-0.06, -0.60)*
Quadratic PS Terms	2.34 *(0.24, 4.40)*	2.88 *(1.04, 4.78)*	-0.37 *(-0.06, -0.68)*	-0.38 *(-0.10, -0.66)*
Cubic PS Terms	2.39 *(2.39, 0.28)*	2.40 *(2.39, 0.37)*	-0.38 *(-0.38, -0.06)*	-0.38 *(-0.38, -0.07)*

Though the results in this model with time from treatment and binary indicator of treatment seem promising in providing a positive treatment effect of PEG, a careful interpretation of the treatment effect estimates is necessary. While there is a significant negative effect on BMI for each unit increase in months of time from treatment to one year post-baseline (-0.34), this effect is additive to the binary treatment indicator at any time. Thus when the time from treatment to 1 year post-baseline is small, there is an overall positive effect of treatment in the measurement of BMI one year post-baseline.

To better visualize the effect of treatment in the time of treatment model,

the difference in mean BMI of treated individuals and untreated individuals is plotted over time from treatment to one year post-baseline in Figure 2.1. As the time between treatment and one year post-baseline increases, the effect of PEG treatment diminishes. One possibility is that this may be due to a waning effect of treatment over time; as outcome is measured further from the time of treatment, the effect of PEG treatment may no longer be discernable.

A naïve analysis, without principal stratification, of the effect of PEG treatment on BMI measured at 1 year post-baseline is provided in Table 2.7. For comparability to SACE and to ensure the measurement of the outcome, only those individuals with survival greater than one year post-baseline are considered in this analysis. Parameter estimates and 95% confidence intervals from a linear regression model are presented. These results indicate that without use of the principal stratification framework, we are unable to identify a significant effect of treatment in this dataset. Both the dichotomous treatment

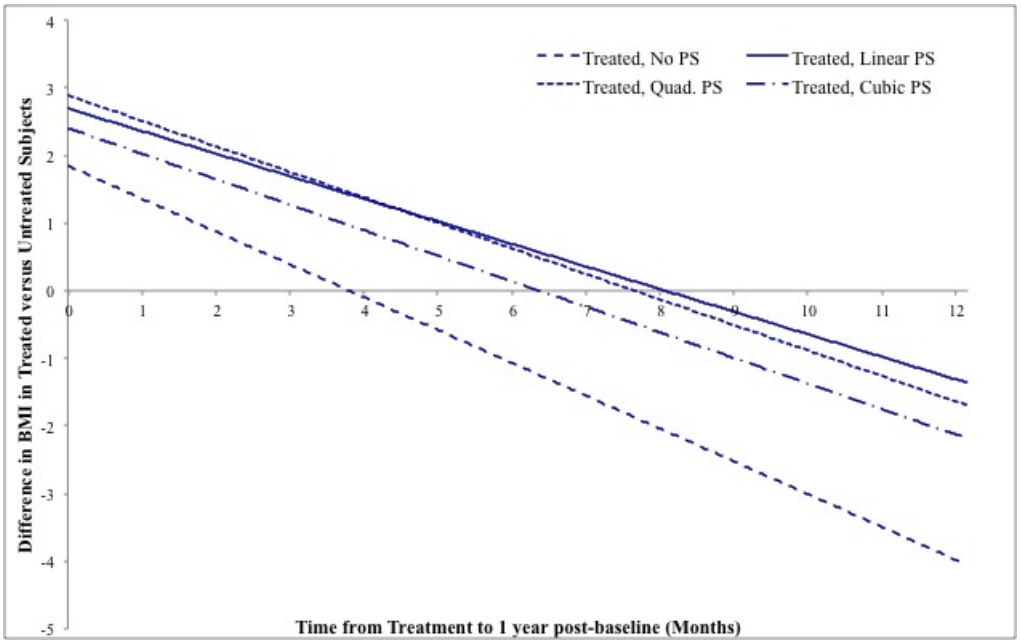

Figure 2.1: Difference in average BMI of treated compared to untreated individuals over time from treatment to 1 year post-baseline (N=491)

Table 2.7: Estimate of PEG treatment effect (with 95% credible intervals) among survivors on BMI measured at 1 year post-baseline without principal stratification (N=278)

	No Propensity Score		Linear PS	
	Mean	*95% CI*	*Mean*	*95% CI*
Model including binary treatment indicator only				
Peg Treatment	-1.38	(-2.62, -0.13)	-0.10	(-1.46, 1.25)
Linear PS Term	-	-	-5.21	(-7.71, -2.71)
Model including both binary indicator and time of treatment				
Time from Treatment to t*	-0.01	(-0.01, 0.00)	0.00	(-0.01, 0.00)
Peg Treatment	-0.07	(-2.28, 2.15)	0.75	(-1.42, 2.92)
Linear PS Term	-	-	-1.48	(-2.11, -0.83)

model and the time of treatment model return non-significant estimates of treatment effect, regardless of the use of propensity scores. These results provide further evidence in support of the use of principal stratification framework in the presence of censoring by death.

2.4 Simulation Studies

We evaluate the proposed methods in Monte Carlo simulations of 500 datasets. Each Monte Carlo dataset contains 500 observations with patient characteristics, principal stratum assignment, treatment information, and outcome data generated as described in the following sections. In all simulation scenarios, principal stratum, treatment assignment, and survival status must be determined prior to outcome generation, as Y_i can only exist for those individuals with $S_i = 1$. Therefore, while Y_i is generated for all observations with $G_i = LL$, it can only be generated for observations with $G_i = LD$ and $Z_i = 1$ or $G_i = DL$ and $Z_i = 0$, and it cannot be generated for observations wtih $G_i = DD$ at all.

In each of the 500 Monte Carlo datasets, four variables are generated to

represent patient characteristics $\mathbf{D}$. $\mathbf{D_1}$, $\mathbf{D_2}$, $\mathbf{D_4}$ are generated from Uniform distributions of varying ranges ($Unif[1,1]$, $Unif[-2,2]$, and $Unif[0,1]$ respectively). $\mathbf{D_3}$ is generated for each individual using a Bernoulli distribution with $p = 0.5$. Overlapping subsets of these covariates are used in the models that generate principal strata assignment ($\mathbf{D_2}$, $\mathbf{D_3}$, $\mathbf{D_4}$) and the treatment assignment ($\mathbf{D_1}$, $\mathbf{D_2}$, $\mathbf{D_3}$) for each observation.

Principal strata assignment, G_i is generated from a discrete distribution with probabilities generated from a multinomial logit model, as presented in equation (2.1) in Section 2.2.3. The parameters of this model, $\boldsymbol{\alpha}_g$, are selected such that LD is the reference group and the average values of the probabilities of principal strata in simulation population have the preferred relationship $\pi_{LL} > \pi_{LD} > \pi_{DL} > \pi_{DD}$. Specifically the parameter values are $\boldsymbol{\alpha}_{LL} = [0.85, 1.0, -0.5, 0.5]$, $\boldsymbol{\alpha}_{DL} = [-0.55, 1.5, -0.5, 0.25]$, and $\boldsymbol{\alpha}_{DD} = [-1.2, -0.8, -0.5, -0.5]$. Treatment assignment, Z_i, is generated from a Bernoulli distribution with probability $p_{Z,i}$. A logistic regression model is used to determine $p_{Z,i}$, with parameters $\boldsymbol{\beta} = [0.1, 1.5, -1.0, -0.5]$, selected to achieve an average probability that is slightly greater than 0.5 for the simulation population.

Knowledge of principal strata and treatment assignments allow for the extrapolation of survival status S_i for each observation. Finally, an outcome measurement is simulated for those observations with a survival status $S_i = 1$ using outcome distributions, $f_g = N(\mathbf{X}_{1,g}\boldsymbol{\eta}_g, \sigma_g^2)$, presented in Section 2.2.4. The design matrix for the LL stratum, $\mathbf{X}_{LL} = (1, \mathbf{Z}, \mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3)$, differs from those of the LD and DL by inclusion of the treatment assignment covariate ($\mathbf{X}_{LD} = X_{DL} = (1, \mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3)$). The outcome model parameters for each stratum are set to $\boldsymbol{\eta}_{LL} = [6, 4, 1.5, 0.8, 0.4]$, $\boldsymbol{\eta}_{LD} = [1, 1, 0.4, 0.2]$, and $\boldsymbol{\eta}_{DL} = [2, 1, 0.4, 0.2]$.

Table 2.8: Results of simulations with binary treatment indicator

Treatment Effect Estimate (LL Stratum)	Relative Bias (%)	Mean SE	Empirical SD	Coverage Probability
Without Monotonicity (All 4 Principal Strata)				
No PS Term	33.63	0.266	0.265	0.00
Linear PS Term	3.78	0.224	0.221	0.93
Quadratic PS Terms	2.17	0.221	0.219	0.92
Cubic PS Terms	1.93	0.225	0.221	0.95
With Monotonicity Assumption (No DL Stratum)				
No PS Term	32.82	0.190	0.190	0.00
Linear PS Term	2.49	0.170	0.170	0.95
Quadratic	1.88	0.171	0.171	0.97
Cubic	2.03	0.172	0.172	0.96

We also consider Monte Carlo simulations in which the monotonicity assumption holds. Three strata are considered, with the he DL stratum removed from the framework. Parameter values $\boldsymbol{\alpha}_g$ for the strata model and $\boldsymbol{\eta}_g$ for the outcome model remain the same for LL, LD, and DD. It is noteworthy that the assumptions made for the data generation in each scenario are imposed on the statistical analysis of the data. That is, when data are simulated with a monotonicity assumption, it is analyzed in the same manner; when data is simulated without the monotonicity assumption, no monotonicity assumption is imposed on the statistical analysis.

The estimated effects of treatment for various dichotomous treatment models are presented in Table 2.8. These results include simulations with and without the monotonicity assumption and the incorporation of higher order propensity scores. The monotonicity assumption reduces the number of principal strata to three by eliminating the DL stratum, with all other models, distributions, and true parameters remaining the same as when the monotonicity assumption is not employed. Propensity score is calculated as the probability of treatment using a logistic regression model for the generated Z_i.

The estimates of effect when a propensity score term is not included in the

model with and without monotonicity assumption exhibit substantial bias. However, the inclusion of propensity scores as linear predictors, either singularly or with higher order terms, reduces the bias of the effect estimates considerably. The relative bias of the effect estimate of treatment is approximately 33% in both scenarios with and without the monotonicity assumption when propensity scores are absent, but this relative bias shrinks to 1.9%-3.8% when propensity score terms are present. Additionally, the coverage probabilities in the presence of propensity scores range from 92%-95% in simulations without the monotonicity assumption and 95%-97% in simulations with the monotonicity assumption. While relative bias is smallest and coverage probability is highest when cubic propensity score terms (and corresponding lower order terms) are included in the simulation framework, those models with linear or quadratic propensity score terms also perform well.

2.5 Discussion

The use of propensity scores within the principal stratification framework allows for the estimation of an unbiased principal effect of treatment, particularly for observational data or randomized data in which the assumption of no unmeasured confounding is suspect. The removal of bias by inclusion of propensity scores in the principal strata and outcome models is evidenced by the results of the simulation studies. It is noteworthy that in order for the principal effect estimate to be unbiased, the assumption of strongly ignorable treatment assignment must hold. In other words, there must be no unmeasured confounders. In the current data analysis from the Emory ALS Clinic, one important confounder that is not available is the Revised ALS Functional Rating Scale (ALSFRS-R) score, a validated instrument for measuring the progres-

sion of ALS. The absence of this confounder could be particularly problematic in achieving ignorability when conditioning on propensity scores. However, this deficiency of confounders in this data would likely render any methods ineffective in establishing balance among treatment groups. In addition, even without the ALSFRS-R score, balance of the observed covariates in Table 2.3 and the change of effect estimates that occurs when the propensity score is included in the model lend support to the validity in our propensity score methods. Future applications of the proposed methods to other data with richer measurements of confounders should further demonstrate the reduction of selection bias and confounding.

Though the effect of the treatment is not significant in the dichotomous treatment model for PEG treatment, a positive and significant effect of treatment is identified in the time of treatment model. As mentioned in earlier sections, it is possible that we are unable to capture the effects of treatment administered at time points further away from the time of outcome measurement due to a waning effect of treatment. Identifying the treatment effect as it changes over time may be of use to the clinical community, requires more sophisticated techniques, and thus is left for future research.

Overall, the results presented from the application to the ALS data are not sensitive to the assumption of monotonicity. In the data application, this may be due to the small proportion of individuals in DL strata when all four strata are considered. When monotonicity is not assumed, most patients are in the LL and DD strata, with a LD and DL strata each comprising less than 5% of individuals each. It is conceivable then that reallocating such a small proportion of individuals when removing the DL stratum would likely not substantially change the effect estimates of the other strata.

Propensity scores are included as linear predictors of the outcome and

principal strata models, which can be quite restrictive when controlling for selection bias and confounding. Adding higher order terms does add some flexibility in the relationships between propensity score and the dependent variables of each of the models as it acts as a polynomial basis regression splines, but other basis functions may be considered in the future for the most effective control for bias.

Future consideration may also be given to jointly modeling the propensity score with the outcome model and principal strata model in the Bayesian framework. This would allow the quantities observed by sampling outcome and principal strata models to affect the posterior of propensity score in each MCMC iteration. While this could provide a more robust propensity score adjustment, Zigler et al. (2013) show that the feedback between model stages in joint modeling can cause biased causal effect estimates if individual covariates are not also adjusted for in the outcome model. This bias should be accounted for if joint modeling of the three models of outcome, propensity scores, and principal strata is proposed.

Chapter 3

Propensity function matching for estimating treatment effect

3.1 Introduction

In observational studies, treatment assignment is often not randomized, resulting in selection bias. The original propensity score methods as proposed by Rosenbaum & Rubin (1983) have been widely used in many settings of statistical analysis to correct for such bias. However, it is relatively unexplored in settings where treatment is administered at various inconsistent times throughout the follow-up period, though many have explored related issues. Zhao & Tsiatis (1997) and Anstrom & Tsiatis (2001) employed inverse propensity score weighting to correct for confounding in observational studies where only the response is right censored. Li et al. (2001), Lu (2005), and Seeger et al. (2005), have examined matching with propensity scores in the presence of varying treatment times, but they only analyzed data with prespecified fixed time intervals for both the administration of treatment and the measurement of time-varying covariates.

In this paper, we address issues that arise when evaluating a treatment effect in an observational study with varying times of treatment and measured time-varying confounders. To accomplish this, we extend the previously established methods for propensity scores, particularly the P-Function of Imai & Van Dyk (2004) and the Generalized Propensity Score (GPS) of Hirano & Imbens (2004). The proposed Propensity Process incorporates and balances the entire covariate profile from baseline until time of interest, allowing for the analysis of observational or non-randomized data, particularly when the data is unstructured such that measurements are not taken at specific time points.

After a brief introduction to the motivating dataset and notation, Section 3.1 is completed with a review of the established methods incorporating the GPS and the P-Function. In Section 3.2, we define the Propensity Process that incorporates covariate processes and establish large sample properties of

this method. In Section 3.3, we present the methods for estimating the GPS and Propensity Process based on observed time-varying covariates and subsequent matching. In Section 3.4, the proposed approach is applied to our motivating data. We conclude with a discussion of strengths and limitations of the proposed methodology in Section 3.5.

3.1.1 Data of the Emory ALS Clinic Registry

Our work is motivated by the data obtained from a registry of patients maintained by the Emory ALS Clinic. Amyotrophic Lateral Sclerosis (ALS) is a rare progressive disorder resulting in the degeneration of both upper motor neurons of the cerebral cortex and lower motor neurons of the spinal cord and peripheral nervous system (Procaccini & Nemergut 2008). In the Emory registry, the median survival post symptom onset is approximately 29 months, noting that there are no curative therapies for ALS (Traxinger et al. 2013). Palliative care options include the surgical insertion of a percutaneous endogastrostomy (PEG) tube to provide enteral nutrition for individuals who are having difficulty swallowing or breathing (Miller et al. 1999).

The data in the ALS registry are measured longitudinally from first clinic visit (baseline) until time of death for each of n individuals, $i = 1, 2, ..., n$. The time of each of these clinic visits is denoted as t_{ij}, where $j = 1...m_i$. Time intervals between visits vary both by individual as well as within individual clinic visits. Outcomes of interest, Y_i, are changes in clinical measurements including body mass index (BMI) and forced vital capacity (FVC) from baseline to a fixed time t^o. The binary indicator of receipt of treatment in the form of a PEG tube insertion prior to t^o is denoted by Z_i. The time of treatment, $T_{Z,i}$, is only observed for subjects who are treated prior to t^o ($Z_i = 1$), otherwise

the subject is not treated with $T_{Z,i} = t^o$, such that $T_{Z,i} \in [0, t^o]$. Finally, let $\mathbf{X}_i$ denote the set of p covariates observed before treatment, which may also be written as the set of two distinct matrices $\mathbf{X}_i = (\mathbf{X}_{1,i}, \mathbf{X}_{2,i}(t_{ij}))$. $\mathbf{X}_{1,i}$ includes p_1 variables measured only at baseline, while $\mathbf{X}_{2,i}(t_{ij})$ includes p_2 variables measured at each clinic visit up to $T_{Z,i}$ ($t_{ij} \leq T_{Z,i}$).

This is a retrospective study, presenting several challenges in the analysis. First, the treatment was not randomized, potentially resulting in selection bias. Second, individual clinic visits were not at fixed times or intervals, and so covariate measurement is generally not available at the time of treatment. The current propensity score methods for non-binary treatment assignment, described in detail in Section 3.1.2, are not directly applicable to account for these complex features of the ALS dataset. Though we may be able to adapt some of these non-binary propensity methods by means of covariate interpolation, the proposed Propensity Process can better address the selection bias and confounding present in the Emory ALS Clinic Registry.

3.1.2 Propensity Score Methods

The estimation of propensity scores and their use in removing bias when treatment assignment is binary has been well established in the literature (Rosenbaum & Rubin 1983, Rubin & Thomas 1996). In these cases, the definition of the propensity score is the probability of receiving a treatment conditional on a set of observed variables $\mathbf{X}$, $b(\mathbf{X}) = P(Z = 1|\mathbf{X})$. However, when the treatment is non-binary, the probability of treatment may either be difficult to quantify or is not the best balancing score. Two options for flexible propensity scores that have been proposed in recent literature include the generalized propensity score (GPS) of Hirano and Imbens (2004) and the

P-Function of Imai and Van Dyk (2004). Both methods are applicable to a wide variety of treatment assignment models, including non-parametric and semi-parametric models.

Generalized Propensity Scores

GPS is defined as $R = r(Z, \mathbf{X})$ where $r(z, \mathbf{x}) = f_{Z|\mathbf{X}}(z|\mathbf{x})$. In words, the GPS is the density function for treatment assignment evaluated at the observed treatment and covariates. In practice, it may be suitable to reduce $\hat{R}_i$ to the linear predictor of a regression model, for ease of computation and analysis. Hirano and Imbens (2004) propose that the GPS satisfies both the properties of a balancing score as well as weak unconfoundedness, the conditional independence of potential outcomes and a single value of treatment given the GPS evaluated at that treatment and observed covariates. This allows the authors to state further that the causal quantities of interests are unbiased when conditional on the GPS. Specifically, the dose response function of treatment is defined as $E[Y(t)] = E_{r(z,\mathbf{X})}\{E[Y(t)|z, r(z, \mathbf{X})]\}$. In addition $\hat{R}_i$ could be used for matching, stratification, or weighting, though these applications are not described in detail by Hirano and Imbens and have been relatively less explored.

In cases such as our motivating data, time to a specific treatment such as PEG, T_Z, is of interest and a Cox Proportional Hazards (PH) model for T_Z, with or without time-dependent covariates, may be utilized for estimating the Generalized Propensity Score. While the formal definition implies that the GPS would be the density function of time to treatment evaluated at a time of interest t, $r(t, \mathbf{x}) = f_{T_Z|\mathbf{X}}(t|\mathbf{x}) = h_{T_Z}(t|\mathbf{x})S_{T_Z}(t|\mathbf{x})$ where $h_{T_Z}(t|\mathbf{x})$ is the hazard function and $S_{T_Z}(t|\mathbf{x})$ is the survival function for T_Z, evaluation of

$h_{T_Z}(t|\mathbf{x})$ is sufficient for the GPS. The time of interest can be baseline ($t = 0$), time of treatment, time of death, or any other time that is suitable for the analysis of interest. Along these lines, Li et al. (2001) and Lu (2005) used the one-dimensional hazard component $h(t|\mathbf{x})$ for propensity score matching where time of treatment was of interest. These studies support the use of the GPS, modeled with time-varying covariates and evaluated at observed time of treatment and covariates, has a balancing property for matched pairs, such that the treatment assignment is independent of covariates within a matched pair or set.

However, both the studies by Li et al. (2001) and Lu (2005) only used the covariates values at the time of interest for matching and did not use the entire longitudinal profiles of the time-varying covariates. In addition, they developed methods to analyze treatment effect within a patient population from a clinical trial in which each individual was followed for up to 4 years post-enrollment with clinic visits every three months for covariate measurement and potential treatment administration. This study design has two implications; first, covariate measurements are available both at the time of treatment administration for treated individuals and at a comparable time point for potential untreated controls. Secondly, those potential controls that are untreated at the time of treatment for a subject in need of a match, would definitely remain untreated until the time of outcome measurement at the next clinic visit. This simplification of the treatment and outcome measurement times allows the use of propensity score methods that are not directly applicable to naturalistic observational studies in which treatment may be administered at any time.

Propensity Function

The propensity function, or the P-Function, is defined as the entire conditional probability density function $e_\psi(\cdot|\mathbf{X}) = p_\psi(\cdot|\mathbf{X})$. Imai and van Dyk suggest using a uniquely parameterized propensity function $\theta_\psi(\mathbf{X})$, such that $e_\psi(\cdot|X)$ depends on X only through θ, as a summary of the conditional density function. Conditional on a calculated P-Function, the probability of treatment assignment is independent of any covariates, $P\{Z = z|\mathbf{X} = \mathbf{x}_1, \theta_\psi(\mathbf{x}_1)\} = P\{Z = z|\mathbf{X} = \mathbf{x}_2, \theta_\psi(\mathbf{x}_2)\}$.

The choice of propensity function is determined by the appropriate regression model for the full conditional distribution of $Z|\mathbf{X}$. The function $\theta_\psi(\mathbf{X})$ is therefore generally determined to be the linear predictor of this regression model. Examples of the characterizing function θ are provided by Imai and van Dyk for several treatment assignment models, including $\theta_\psi(\mathbf{X}) = \beta\mathbf{X}$ for a continuous treatment that follows a normal conditional distribution and $\theta_\psi(\mathbf{X}) = \pi(\mathbf{X})$ for a categorical treatment that follows a multinomial conditional distribution.

Once estimated by regression, the model coefficients inform an estimate of θ. Imai and van Dyk suggest matching or subclassification based on these estimated values, $\hat{\theta}$, to determine the causal effect denoted by ϕ. Additionally, the authors suggest that treatment effect estimate within each subclass may vary smoothly as a function of θ. To accomplish this, a smooth coefficient model, one that allows ϕ to vary with θ, may be fit.

While the GPS as described by Hirano & Imbens (2004) and the P-Function defined by Imai & Van Dyk (2004) are able to include time-varying covariates in the conditional model for treatment assignment, the evaluation of these propensity score methods require evaluation at observed values of treatment and covariates. However, data from observational registries, such as the ALS

dataset, seldom have measurements at similar time points for all subjects post baseline, and thus we propose the time-dependent Propensity Process that balances the entire covariate process from baseline until the time point of interest.

3.2 Time-varying Propensity Process

Following the notation in Section 1, we denote the set of the potential outcomes for subject i by $\{Y_i(t_{Z,i}), t_{Z,i} \in \mathcal{T}\}$, where $\mathcal{T}$ denotes the set of potential treatment times. In our framework for assessing treatment effect, we make two assumptions. First, we assume that the distributions of the potential outcomes for different subjects are independent of each other given $\mathbf{X}$, the covariates observed prior to time of treatment. This is also known as SUTVA or the stable unit treatment value assumption (Rubin 1980, 1990). Second, we assume weak unconfoundedness, or that for all $t_Z \in T_Z$, the treatment time t_Z is independent of the set of the potential outcomes, $Y(t_Z)$, given $\mathbf{X}$, the observed covariates prior to time of treatment. In other words, we must assume the mechanism for time to treatment assignment, must be independent of the outcome observed with a specific time to treatment given the covariates. These assumptions are the same as those made for GPS or P-Functions when time to treatment is considered for the treatment assignment models (Hirano & Imbens 2004, Imai & Van Dyk 2004).

To illustrate our ideas, we consider the case where the time to treatment T_Z is assumed to follow a Cox PH model,

$$h_{T_Z}(t|\mathbf{X}(t)) = h_0(t)\exp\{\boldsymbol{\beta}_1^T\mathbf{X}_1 + \boldsymbol{\beta}_2^T\mathbf{X}_2(t)\}, \tag{3.1}$$

where $h_0(t)$ is the baseline hazard function independent of the covariates. It follows that the density function for T_Z, which defines the treatment assignment model, can be written as

$$f_{T_Z}(t|\mathbf{X}(t)) = h_{T_Z}(t|\mathbf{X}(t))S_{T_Z}(t|\mathbf{X}(t)) = h_{T_Z}(t|\mathbf{X}(t))\exp\{-\int_0^t h_{T_Z}(u|\mathbf{X}(u))du\} \tag{3.2}$$

We define the time-varying *Propensity Process* as $\theta(t) = \boldsymbol{\beta}_1^T\mathbf{X}_1 + \boldsymbol{\beta}_2^T\mathbf{X}_2(t)$, $t \in [0, t^o)$, noting that $\theta(t)$ uniquely determines the density function $f_{T_Z}(t)$ and hence the treatment assignment mechanism. Though at first glance, this concept may be similar to the P-Function, which provides balance on an entire conditional probability density of treatment, the distinguishing factor of the Propensity Process is that $\theta(t)$ depends on t and is of infinite dimension, whereas the P-Function is defined as θ and is of finite dimension. Along the lines of Imai & Van Dyk (2004) and Rosenbaum & Rubin (1983), we establish the theoretical results of the time-varying Propensity Process assuming that $\theta(\cdot)$ is known.

Proposition 1 $T_Z \perp \{\mathbf{X}_1, \overline{\mathbf{X}}_2\} \mid \overline{\theta}$, *where* $\overline{\mathbf{X}}_2 \equiv \{\mathbf{X}_2(t); t \in [0, t^o)\}$ *is the entire time-varying covariate process if treatment is not given in* $[0, t^o)$ *and* $\overline{\theta} \equiv \{\theta(t); t \in [0, t^o)\}$ *is the entire Propensity Process.*

Proposition 1 essentially establishes $\overline{\theta}$ as a balancing score. However, Proposition 1 requires that $\overline{\theta}$ is known or can be estimated in the entire domain $[0, t^o)$. In practice, however, we only are able to observe the covariate process $\mathbf{X}_2(\cdot)$ up to the time of treatment receipt for each subject. The next proposition establishes the covariate balancing property for a given time point t^* in $[0, t^o)$.

Proposition 2 *$T_Z(t^*) \perp \left\{\mathbf{X}_1, \overline{\mathbf{X}}_2(t^*)\right\} \mid \overline{\theta}(t^*)$, where the treatment assignment variable $T_Z(t^*)$ is defined as $T_Z(t^*) = T_Z$ if $T_Z \leq t^*$ and $T_Z(t^*) = N$ if $T_Z > t^*$, $\overline{\mathbf{X}}_2(t^*) \equiv \{\mathbf{X}_2(t); t \in [0, t^*]\}$ is the covariate process up to time point t^*, and $\overline{\theta}(t^*) \equiv \{\theta(t); t \in [0, t^*]\}$ is the Propensity Process up to time point t^*.*

Several remarks are in order. First, while we define the Propensity Process based on the Cox PH model with time-varying covariates, $\theta(\cdot)$ can also be defined along similar lines based on other models for the treatment assignment mechanism that involve time-varying covariates or time-varying coefficients. Second, Proposition 2 provides the theoretical justification for the approach of matching a subject treated at t^* with a subject in the corresponding risk set based on the Propensity Process estimated up to the treatment time t^* for the treated subject, where the risk set is defined as the set of subjects that are treated at a later time or never treated before t^o (i.e., $T_Z > t^*$). It follows that each matched pair will have the same distribution for the covariate process up to t^*. Third, our result is similar in spirit to Proposition 1 in Lu (2005) but is more general in the sense that it balances the entire covariate process up to t^* not just the covariates measured at t^*. A proof for Propositions 1 and 2 is provided in the Appendix.

3.3 Propensity Process: Estimation and Matching

In practice, the Propensity Process must be estimated from the observed data. The challenge for estimating the Propensity Process is that we do not observe the complete covariate process $\mathbf{X}_2(\cdot)$. We only get to observe $\mathbf{X}_2(\cdot)$ at a set of discrete time points for each subject. To fix ideas, we first describe how to estimate the GPS using baseline variables $\mathbf{X}_1$ and then investigate how to estimate the Propensity Process as well as the GPS using both baseline

variables and time-varying variables $\mathbf{X}_2$. For assessing the treatment effect, we focus on the approach of matching a treated subject with a subject in the corresponding risk set based on the estimated GPS or Propensity Process, followed by conducting hypothesis testing using matched pairs. The definition of the risk set is provided in Section 3.3.3.

3.3.1 Generalized Propensity Score Using Baseline Variables

First, a Cox PH model that includes covariates measured at baseline only $(\mathbf{X}_{1i}, \mathbf{X}_{2,i}(0))$ is considered, as shown in equation (3.3).

$$h_{BL}(t|\mathbf{X}) = h_0(t)exp\left(\boldsymbol{\beta}_{BL,1}^T\mathbf{X}_1 + \boldsymbol{\beta}_{BL,2}^T\mathbf{X}_2(0)\right) \tag{3.3}$$

We denote the parameter estimates by $\widehat{\boldsymbol{\beta}}_{BL,1}$ and $\widehat{\boldsymbol{\beta}}_{BL,2}$. The linear predictor of this model is the estimated propensity score, $GPS_{BL,i} = \widehat{\boldsymbol{\beta}}_{BL,1}^T\mathbf{X}_{1,i} + \widehat{\boldsymbol{\beta}}_{BL,2}^T\mathbf{X}_{2,i}(0)$, and is used to determine the distance between individuals for matching. Since the GPS is a scalar, we can use standard distance metrics such as the squared linear distance, $Q_{m,BL} = (GPS_{BL,i} - GPS_{BL,i'})^2$, where individual i would be treated and individual i' would be in the corresponding risk set.

Though this method presents an opportunity for a straightforward calculation of the GPS and a simple metric for matching, the treatment assignment model is only able to reduce or remove potential selection bias between treatment groups due to unbalanced baseline variables. If treatment assignment was decided at first clinic visit, this method would be sufficient. However, the decision for treatment is generally made after baseline in this population and likely corresponds to disease progression, implying it is necessary to control for the treatment assignment mechanism that depends on time-varying covariates.

3.3.2 Time-Varying GPS and Propensity Process

Inclusion of time-varying covariates in the Cox PH model allows for the evaluation of the GPS at the time of treatment and the entire Propensity Process from the baseline to the time of treatment, and hence the removal of selection bias due to unbalanced time-varying covariates. The model, presented in equation (3.4), incorporates both the covariates that are measured at baseline and remain static over time ($\mathbf{X}_1$), as well as those that have repeated measurements over time ($\mathbf{X}_2(t)$).

$$h_{TV}(t|\mathbf{X}(t)) = h_0(t)exp\left(\boldsymbol{\beta}_{TV,1}^T\mathbf{X}_1 + \boldsymbol{\beta}_{TV,2}^T\mathbf{X}_2(t)\right). \tag{3.4}$$

We denote the parameter estimates by $\widehat{\boldsymbol{\beta}}_{TV,1}$ and $\widehat{\boldsymbol{\beta}}_{TV,2}$. We use model (3.4) with time-dependent covariates to define the estimated GPS and the estimated Propensity Process for matching. For subject i, the estimated GPS is defined as the the linear predictor of this model evaluated at the time of treatment, i.e., $GPS_{IP,i} = \widehat{\boldsymbol{\beta}}_{TV,1}^T\mathbf{X}_1 + \widehat{\boldsymbol{\beta}}_{TV,2}^T\mathbf{X}_2(T_{Z,i})$ and the estimated Propensity Process $PP_i = \{\hat{\theta}_i(t) = \widehat{\boldsymbol{\beta}}_{TV,1}^T\mathbf{X}_1 + \widehat{\boldsymbol{\beta}}_{TV,2}^T\mathbf{X}_2(t); t \in [0, T_{Z,i}]\}$, is the full functional process from baseline until time of treatment. Computation of these propensity scores is not straightforward. First, the time-dependent covariates $\mathbf{X}_2(\cdot)$ are often not measured at time of treatment for treated individuals. Secondly, untreated individuals within the risk set are unlikely to have covariate measurements at the exact time of treatment for their potential treated match. Thus, within the matching process, all individuals are subject to interpolation of the covariate value at a specific time by a model that estimates the covariate process over time.

To estimate or interpolate the covariate process, we model each time-

dependent covariate using a mixed model with both fixed and random effects of the time variable. This allows a predictive curve to be fit for the covariate process of each individual from baseline until the end of follow-up, thereby providing values for each covariate over the continuous support of the follow up time. For individual i ($i = 1, \ldots, n$) at times t_{ij} ($j = 1, \ldots, m_i$), the model for covariate k ($k = 1, \ldots, p_2$) in $\mathbf{X}_{2,i}$ is given in equation (3.5).

$$X_{2,i}^{(k)}(t_{ij}) = \mu^{(k)}(t_{ij}) + \nu_i^{(k)}(t_{ij}) + \epsilon_{ii}^{(k)}, \tag{3.5}$$

where $\mu^{(k)}(\cdot)$ is the population average nonparametric curve (i.e., the fixed effect), $\nu_i^{(k)}(\cdot)$ is the subject-specific curve (i.e., the random effect), and $\epsilon_{ij}^{(k)}$ are independent random errors. To estimate $\mu^{(k)}(\cdot)$ and $\nu_i^{(k)}(\cdot)$, we can use the approach of basis expansion, that is, write $\mu^{(k)}(\cdot) = \mathbf{b}(\cdot)^T\boldsymbol{\gamma}^{(k)}$ and $\nu_i^{(k)}(\cdot) = \mathbf{b}(\cdot)^T\boldsymbol{\alpha}_i^{(k)}$, where $\mathbf{b}(\cdot)$ is a set of basis functions such as polynomial basis or cubic spline basis and $\boldsymbol{\gamma}^{(k)}$ and $\boldsymbol{\alpha}_i^{(k)}$ are two sets of coefficients for the basis functions. Using the estimated $\mu^{(k)}(\cdot)$ and $\nu_i^{(k)}(\cdot)$, the values of each time-varying covariate may be interpolated from the model at any time of interest from baseline until death, and then can be used to determine the value of the GPS_{IP} and the time-dependent PP.

Figure 3.1 provides an example of the estimated curves of 4 examples of observed time-varying covariate values over time from baseline until t^o. In each scatter plot, blue dots indicate an observed covariate measurement ($\mathbf{X}_{2,i}(t_{ij})$) at various times of measurement (t_{ij}) for each individual. Red dots are covariate measurements at time of treatment, which are interpolated from the estimated curve for each individual. The blue lines in each plot represent the individual curves fit by model (3.5).

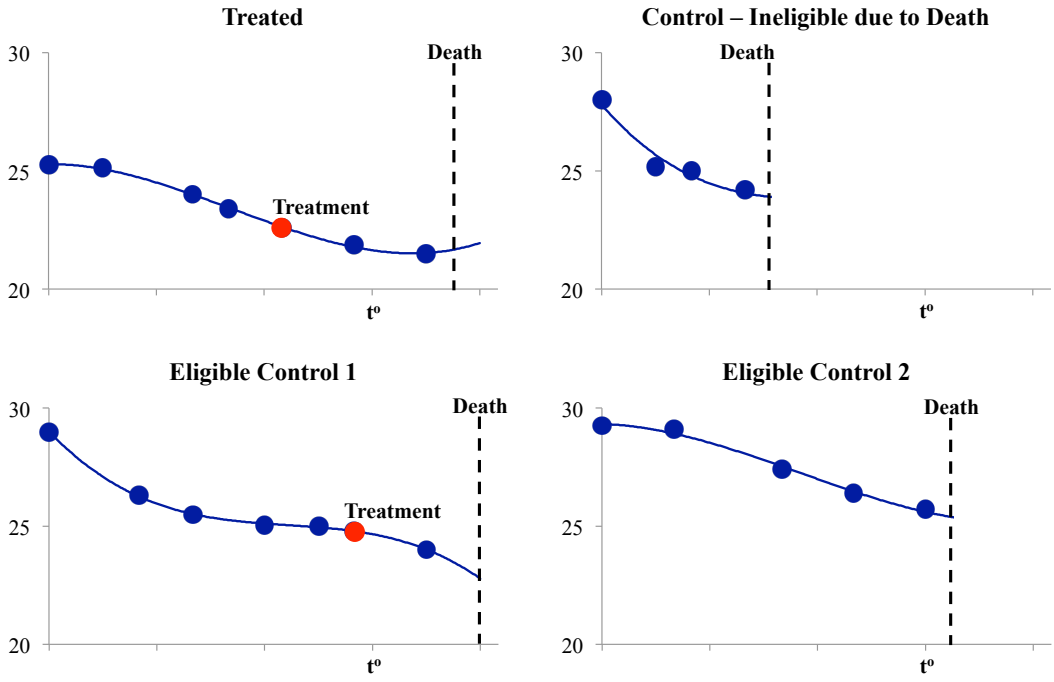

Figure 3.1: Examples of observed time-varying covariate processes

3.3.3 Definition of Risk Sets and Matching

Matching is performed by minimizing the squared distance between the values of the GPS_{IP} or PP of treated and individuals in the corresponding risk set. The GPS_{IP} is a scalar value for each individual, the squared distance between a treated individual i and a member of the risk set i' is computed in the same manner as in the case of GPS_{BL}, $Q_{m,IP} = (GPS_{IP,i} - GPS_{IP,i'})^2$. On the other hand, since the PP is a function of time for each individual, we define the metric for matching in equation (3.6), which is the integration of the squared difference of the two curves with respect to time.

$$Q_{m,PP} = \int_0^{T_{Z,i}} \left(PP_i(t) - PP_{i'}(t)\right)^2 dt \tag{3.6}$$

The risk set of a treated individual i includes both individuals who never receive treatment during follow up and those who receive treatment later than

the treatment time of individual i, i.e., $T_{Z,i'} > T_{Z,i}$. Referring once again to Figure 3.1, "Treated" represents an individual who is treated at T_Z, its risk set for matching may include Eligible Controls 1 and 2 as $T_{Z,EC1} > T_Z$ and $T_{Z,EC2} = t^o$ if they have not been matched to another treated individual but the individual labeled as "Control-Ineligible due to Death" is excluded from the risk set as $T_{S,IC} < T_Z$.

Risk sets of individuals who are available for matching are built chronologically by time using sequential matching algorithms. Individuals in the risk set with the smallest distance Q_m are matched to the individual with the earliest time of treatment, $T_{Z(1)}$. After matching, all members of matched pairs must be removed from risk sets of later time points, and the process is continued chronologically by $T_{Z,i}$ until all treated individuals are matched or until there are no suitable controls available for matching.

3.3.4 Hypothesis Testing

The difference in the outcome of interest is calculated for each matched pair m (m= 1,2,...,M) as $\Delta Y_m = Y_{m,i} - Y_{m,i'}$. The Wilcoxon signed rank test is used to test whether the difference across matched pairs is different from 0. The null hypothesis of this test is $H_0 : Y(T_Z) = Y(T'_Z)\ \forall\ T_Z < T'_Z$, i.e., the outcomes are the same for individuals who are treated at some time t and those individuals who are untreated at that time t, regardless of whether the individual is later treated or never treated.

3.4 Analysis of Data from ALS Registry

The data from the Emory ALS Clinic Registry includes 697 individuals who have died before July 31, 2011, with their dates of death cross-referenced with the Social Security database. Those individuals in the dataset who received PEG treatment, did so after their first visit to the Emory ALS Clinic. The treatment of surgical insertion of the PEG tube is offered to individuals to supplement or be a primary source of nutrition, but the timing of the recommendation by the physician involves many factors and the final decision to have the surgical insertion of the tube is made by each patient. Though there is little conclusive or corroborated evidence of a significant positive benefit of PEG, many neurologists have noted anecdotal evidence from their medical practice that patients with PEG fare better than those without (Gelinas & Miller 2000). One possibility is that assessment of the effect of PEG has been complicated by issues such as non-randomized treatment receipt that arose in an observational setting such as the Emory ALS Registry.

We consider two outcomes of interest in the ALS registry data. Both the change in body mass index (BMI) and the change in forced vital capacity (FVC) from baseline to 24 months post baseline are tested for association with PEG treatment. In shorthand, these outcomes are denoted as ΔBMI_{24} and ΔFVC_{24} respectively. All observed covariates are included in the estimation of the Cox proportional hazards treatment assignment models for each of the propensity scores. The baseline risk factors $\mathbf{X}_{1,i}$ include age at diagnosis (DX), sex, site of onset of disease, negative inspiratory force (NIF), and time from diagnosis to the first clinic visit (CV). The two time-varying measurements that may be included in $\mathbf{X}_{2,i}(t_{ij})$ are also the outcomes of interest, BMI and FVC. Therefore, when BMI is considered to be the outcome of interest, FVC is included in as a covariate in the propensity score models, and likewise BMI is

covariate in the propensity score models when FVC is the outcome of interest. It is noteworthy that these covariates may not be measured at every clinic visit for every individual, with reasons for missingness varying by individual and clinic visit. Each time-varying covariate is modeled over time with a linear mixed model, which includes population average cubic (and lower order) terms for time and a subject-specific linear term for time. The estimated curves are used to interpolate the covariate values needed for matching.

We test for an association of PEG treatment on the outcomes of BMI and FVC in four scenarios. First, a naïve analysis compares all treated individuals to those who are untreated prior to t^o. The additional three scenarios include matched analyses using the three propensity score methods described in Section 3.3: Baseline GPS, Interpolated GPS, and Propensity Process. In each matched analysis, a match threshold ω is employed and may be specific to the propensity score method.

Prior to matching, 396 individuals who have survived 24 months post baseline have measurements of BMI both at baseline and a time within 6 months of the t^o. Matching results in 95 pairs when using the Baseline GPS, 94 pairs when using the Interpolated GPS, and 98 pairs when using the Propensity Process. In addition, 441 individuals have survived 24 months post baseline

Table 3.1: Balance of covariates as measured by p-values of covariate effect in the time-dependent Cox PH model for time to treatment stratified by matched pairs for FVC outcome model

Covariate	Non-stratified Cox PH	BL GPS Stratified	IP GPS Match Stratified	PP Match Stratified
BMI	0.716	0.074	0.972	0.723
NIF	0.892	0.400	0.323	0.478
Age at DX	0.016	0.000	0.559	0.752
Sex	0.244	0.912	0.324	0.454
Site of Onset	0.030	0.005	1.000	0.891
Time from DX	0.463	0.235	0.042	0.431

and have measurements of FVC both at baseline and a time within 6 months of the t^o. After matching, there are 98 pairs when using the Baseline GPS, 97 pairs when using the Interpolated GPS, and 93 pairs when using the Propensity Process.

Following the example of Li et al. (2001) and Lu (2005), balance of covariates is examined by examining p-values from a Cox proportional hazards model for time to treatment with each of the covariates in $\mathbf{X}$ included one at a time. In the matched populations, the model is stratified by the m matched pairs. The results of checking balance with the propensity score methods including time-dependent covariates are available in Tables 3.1 and 3.2. Prior to matching, many covariates are unbalanced amongst the treated individuals and the populations that are considered to be their controls. While matching using the Baseline GPS and Interpolated GPS does seem to help balance some covariates across the treatment groups, there are still a few covariates for which balance is not achieved. Baseline GPS matching does not balance age at diagnosis or site for when used for either the outcome model of BMI or FVC. Interpolated GPS does not balance the time from diagnosis to first clinic visit when used for the outcome model of FVC only. However, matching using the Propensity Process results in balance across all covariates regardless of

Table 3.2: Balance of covariates as measured by p-values of covariate effect in the time-dependent Cox PH model for time to treatment stratified by matched pairs for BMI outcome model

Covariate	Non-stratified Cox PH	BL GPS Stratified	IP GPS Match Stratified	PP Match Stratified
FVC	0.019	0.264	0.152	0.139
NIF	0.892	0.548	0.520	0.524
Age at DX	0.016	0.014	0.170	0.885
Sex	0.244	0.150	0.338	0.904
Site of Onset	0.030	0.003	0.793	0.432
Time from DX	0.463	0.237	0.618	0.301

the outcome model considered. This indicates matching with the Propensity Process outperforms the matching by Baseline or Interpolated GPS in terms of balancing covariates.

Table 3.3 presents hypothesis testing for the effect of PEG on change in BMI at 24 months from baseline and change in FVC at 24 months from baseline. Of note, in the naïve analysis, the Wilcoxon rank sum test is used for hypothesis testing, rather than the Wilcoxon signed rank test which is used for all matched analyses. It is also important to note that because the naïve analysis computes the difference in outcome for treated individuals compared to strictly untreated individuals, one should exercise some caution when directly comparing the results of this method to the matched analyses.

Overall, the results of the Propensity Process matched analysis are quite different from the results of the naïve analysis and suggest that there is a positive or protective effect of treatment on BMI and FVC. In fact, when comparing those individuals who have been treated to those who have never been treated or later treated that have been matched using the Propensity Process, treatment is significantly associated with positive change in BMI from baseline to 24 months. This is markedly different than the naïve analysis results of a significant negative change in BMI from baseline to 24 months. There is no significant effect of PEG on the change in FVC from baseline to 24 months after any propensity score matching. However, the direction of the median

Table 3.3: Wilcoxon test for the median difference in change in FVC and BMI from baseline to 24 months

	Median Diff. in ΔFVC_{24}	Test p-value	Median Diff. in ΔBMI_{24}	Test p-value
Naïve	-0.42	0.006	-14.72	<0.001
Baseline GPS	-0.72	0.123	-7.13	0.060
Interpolated GPS	0.18	0.228	1.58	0.551
Propensity Process	0.48	0.042	2.07	0.456

difference is positive after matching by Propensity Process, in contrast to a significantly negative median difference in the naïve analysis. It is possible that with a larger sample size or additional measured confounders, Propensity Process matching may return a significant positive effect of treatment on the change in FVC from baseline to 24 months.

Finally, neither the results of Baseline GPS matching or Interpolated GPS matching return a significant effect of treatment on the change in either outcome from baseline to 24 months. However, the directions of the median matched differences of Baseline GPS matching and Interpolated GPS matching are the same as those of the naive analysis and the Propensity Process matching respectively. This, in combination with the results from testing for balance may indicate that the Interpolated GPS performs better than the Baseline GPS, though not as well as the Propensity Process.

3.5 Discussion

The Propensity Process offers the advantages of balancing time-varying covariates over the observed covariate process from baseline to time of treatment. Without modification such as the interpolation of covariate processes, other propensity score methods may restrict balance of treatment groups to a single time of observed treatment and covariates. Therefore, matching using the Propensity Process removes selection bias and confounding when the data is not structured in a manner that allows for use of the established propensity score methods. A key component to this process is the interpolation of covariate curves. While in this example we use long-linear mixed models with splines to ensure flexibility in the model and parametric assumptions, there must also be enough individual longitudinal data collected to estimate these

curves. This may be a limitation in data with sparsely collected longitudinal covariates.

A matched analysis is used for the Emory ALS clinic data as a straightforward method of hypothesis testing when time of treatment varies by individual. Analysis of treatment effect among propensity score matched pairs has long been considered a method for removing bias. There may exist a data structure in which modeling or a weighted analysis may be possible for the estimation of a causal effect using the Interpolated GPS or Propensity Process. However, for data such as the Emory ALS clinic registry, implementation of weighted analyses or modeling would be difficult. The primary reason for this is that while the choice of a time of evaluation of the propensity score methods may be natural for treated subjects, a complimentary or comparable time frame for untreated population would be unclear. Matched analyses allow us to sidestep this issue by matching control individuals to treated individuals at a given time of treatment.

Matching by any type of propensity score is subject to the limitations of the propensity score models themselves. In particular, if there are unmeasured confounders that are important for the treatment assignment model or for treatment effect on outcome, we cannot be confident that our matching will remove selection bias or confounding as our propensity score methods will not be sufficient. Additionally, the propensity score matching relies on the correctly specified model for treatment assignment. A misspecified model would result in a poorly matched analysis and unbalanced treatment groups. Finally, the probability of treatment must be bounded away from 0 or 1 to avoid issues caused by a priori counterfactual groups. Though it may be reasonable to assume that the model for time to PEG treatment is correctly specified and that the probability of treatment is bounded away from 0 and 1 in the Emory ALS

data, the existence of unmeasured confounders is likely. In particular, there are measures of ALS disease progress, such as the revised ALS Functional Rating Scale (ALSFRS-R) score, that are not available in the observed covariates in this data, but may ensure the removal of bias if included in propensity score methods. However, given the balance achieved by the Propensity Process matched analyses, it is possible that by controlling for covariate processes over time we are able to sufficiently remove the time-varying bias and confounding.

Another potential limitation of the analysis is the exclusion of individuals who die prior to t^o. By excluding these individuals, we are able to avoid complications of censoring by death, described in detail by Rubin et al. (2006) and Frangakis et al. (2007). Future extensions of these methods could address censoring by death, such that no exclusions are necessary.

Chapter 4

Estimating the palliative effect of percutaneous endoscopic gastrostomy in an observational registry in the presence of missing outcome data

4.1 Introduction

Though observational data from disease registries and clinics may offer an opportunity to examine the effect of treatment when clinical trials may not be plausible, these data are often fraught with analytical challenges. This is particularly true of a disease with high disability and mortality rates as issues of unmeasured data may occur due to a patient's absence or death. Thus the selection of a single time point or even a range of times post-baseline for outcome measurement results in missing outcome measurement.

In this paper, we are interested in estimating a causal effect of treatment in a clinical registry of patients with amyotrophic lateral sclerosis (ALS), a neurodegenerative disorder with a very poor prognosis, but there are several characteristics of the data that require consideration in the analysis. (Gelinas & Miller 2000, Procaccini & Nemergut 2008). The most obvious obstacle to causal inference in this data is the non-random treatment assignment, which may lead to issues of selection bias and confounding. Secondly, the data are collected longitudinally for each individual at their chosen times of clinic visits. This means that there are varying times elapsed between each individual's clinic visits and amongst all individuals. This can be especially problematic in determining a time for outcome measurement, as those individuals who do not have a clinic visit during the measurement time, and therefore do not have an observed outcome, may be categorically different from those who do. Finally, censoring by death of the outcome may be caused by the fatal and fast-progressing nature of ALS. All three of these issues must be addressed in the framework for analysis for an unbiased estimation of treatment effect.

The data in the ALS registry is measured longitudinally at patient clinic visits occurring at uneven intervals of time for each individual, $i = 1, 2, ..., N$. The baseline visit is defined as the first clinic visit for all patients and is de-

noted by t_0. The set of patient characteristics, denoted by $\mathbf{D}$, includes both a set of characteristics measured only at baseline and a set measured at each clinic visit. Survival time post-baseline, T_S, is observed for all individuals in the clinic registry, and has been confirmed by Social Security database records.

The surgical insertion of a percutaneous endoscopic gastrostomy (PEG), a palliative procedure that provides enteral nutrition, may be administered at any time post baseline. Those individuals who receive PEG prior to the time of outcome measurement, t^o, are considered treated, regardless of their survival until this time point. Treatment variables include both as time to treatment, T_Z, as the binary indicator of treatment, Z. Untreated individuals have values $Z = 0$ and $T_Z = T_S$ for those individuals who do not survive until t^o, or $Z = 0$ and $T_Z = t^o$ for those who survive past t^o. Treatment status is known for all individuals in the clinic registry at all times during follow-up.

Of scientific interest is the effect of this palliative PEG procedure on an outcome of body mass index (BMI) measured at a specific point post baseline. The outcome is denoted as $\mathbf{Y}$, and is observed if an individual survives until measurement time and if clinic visit occurs within the range of $t^o \pm \delta t$, where δt is some short time period for outcome measurement. The outcomes of those individuals who survive until the time of outcome measurement but do not have a clinic visit within $t^o \pm \delta t$ are considered missing, and therefore have a missing indicator value $M = 1$. Those individuals who do not survive until the time of outcome measurement have values of Y and M that are undefined and can be considered "censored" by death. Figure 4.1 illustrates the potential outcomes for all individuals.

In the context of Rubin's Causal Model, principal stratification offers a framework for causal inference in the presence of a complicating post-treatment variable. This principal stratification framework is described in detail by Fran-

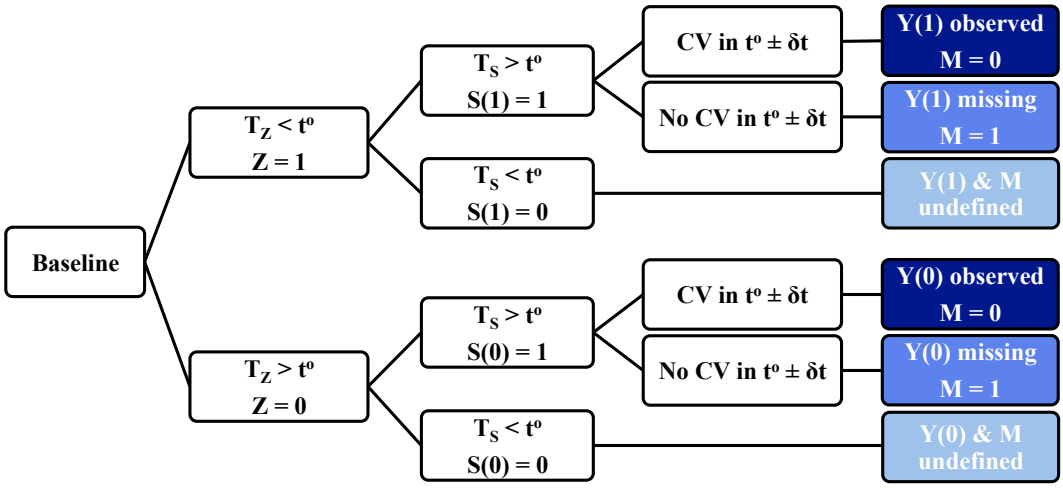

Figure 4.1: Potential outcomes for all patients given survival, treatment, and clinic visit time.

gakis & Rubin (2002*a*), and Zhang & Rubin (2003), who describe the appropriate methodology when a post-treatment variable such as survival or graduation censors the outcome of interest. Zhang et al. (2009) further outlines specific parametric approaches for the identification of survivor average causal effect (SACE) within principal strata. In these examples, however, a randomized and prospective study design assures the complete collection of outcome data, an advantage not guaranteed in our retrospective clinic registry data.

To address missing outcome data, Frumento et al. (2012) simultaneously address post-treatment variables and unobserved outcomes within a principal stratification framework. The authors suggest that if the missing mechanism of the outcome data is non-ignorable, a suitable assumption for the missingness within a principal stratification framework is latent ignorability. Latent ignorability indicates that if principal strata membership is known for each individual, the missing mechanism is ignorable. The authors posit two sets of exclusion restriction assumptions that constrain the probabilities of missingness for individuals for given values of treatment assignment and post-treatment variables. When principal strata membership is unknown under latent ignor-

ability, the missing mechanism is not ignorable and must be included in the modeling framework, regarless of the exclusion restrictions employed.

It is noteworthy, though, that the motivating dataset for Frumento et al (2012) was a randomized design, and so the authors did not have to account for the complications of an observational study. This is particularly important, because the unbiased estimation of a principal effect, defined as the causal effect within principal strata, relies on an assumption of ignorability or no unmeasured confounding (Frangakis & Rubin 2002*a*). When selection bias or confounding may be present, either as residual confounding in a randomized clinical trial or due to observational data, Schwartz et al. (2012) show that the resulting principal effect estimate is likely to be biased. This result indicates that in the absence of randomization or when the randomization scheme results in poor balance among treatment groups, there is a need to incorporate methods for alleviating selection bias and confounding within a principal stratification framework.

Of the many methods that address selection bias or confounding in observational studies, a popular choice is the propensity score introduced by Rosenbaum & Rubin (1983). The propensity score provides a means of balancing covariates across treatment groups, a result that would otherwise be guaranteed if a randomized study design is used. Though propensity scores were introduced for balancing treatment assignment groups when treatment is binary, other authors have extended these methods to non-binary treatment assignment models, such as generalized propensity score methods (Imai & Van Dyk 2004, Hirano & Imbens 2004). These methods allow for the estimation of causal effects when treatment assignment is ordinal, categorical, continuous, semi-continuous, or even multivariate. In the proposed methodology, the generalized propensity score is estimated using a proportional-hazards

model for time to treatment.

The methodology proposed in this paper incorporates generalized propensity scores and consideration for missing outcomes in a principal stratification framework. Jo & Stuart (2009) include propensity scores within a principal stratification framework, not for the removal of selection bias or confounding as the data was randomized, but instead to predict principal strata membership in a matched analysis. Thus the complications of using a principal stratification framework in data from an observational registry. Furthermore, propensity score methods for non-binary treatment assignment have not yet been employed for conditional ignorability in a principal stratification framework when missing outcomes are also a concern in any type of study. We present methods for the practical application of a principal stratification framework in data further complicated by issues of selection bias and missing outcomes.

4.2 Methodology

Patient characteristics in combination with time to treatment, indicator of no treatment, and the propensity score vector **PS**, as described in the following section, comprise the matrix of observed data **X**, which is used in parts for modeling. All patient characteristics **D** are multiply imputed at baseline. In the regression model for the outcome **Y**, the subset of **X** that is included in the analysis is $\mathbf{X}_1$, which may include $\mathbf{T_Z}$, **Z**, **PS**, and a subset of variables from **D**. The inclusion of treatment variable depends on the outcome model and the stratum, as described in later sections. $\mathbf{X}_2$ and $\mathbf{X}_3$ are the subsets of **X** included in the regression model for the principal strata probabilities, $P(G = g)$, and in the regression model for missing outcomes, $P(M = 1)$, respectively.

4.2.1 Framework for Causal Inference

Two general assumptions are employed for the causal inference framework. First, we make the Stable Unit Treatment Value Assumption (SUTVA) as defined by Cox in 1958 and summarized by Rubin (1980). This assumption states that there is no interference among the potential outcomes of one individual and the treatment choices of another individual. Second, we assume Strong Ignorability of Treatment Assignment (Rosenbaum & Rubin 1983), which states the distribution of the potential outcomes is independent of treatment assignment, given the observed covariates. The ignorability assumption often proves to be non-trivial, particularly for data from observational studies.

As earlier defined, the indicator for treatment from baseline until t^o (time of outcome measurement) is $\mathbf{Z}$ and the outcome of interest is $\mathbf{Y}$. A post-treatment indicator of survival past the t^o is $\mathbf{S}$. Using the Rubin Causal Model (Holland 1986) as a framework for causal inference, we can define potential outcomes $\mathcal{Y} = \{Y_i(z^P), z^P \in \mathcal{Z} \text{ for } i = 1...n\}$ and $\mathcal{S} = \{S_i(z^P), z^P \in \mathcal{Z} \text{ for } i = 1...n\}$, where $\mathcal{Z}$ is the set of potential treatment values and $Y_i(z^P)$ and $S_i(z^P)$ are the potential outcomes for a given potential treatment z^P.

4.2.2 Generalized Propensity Scores

When considering only the effect of a dichotomous treatment, standard propensity scores may be employed. However, when treatment is defined as a measure of time, we must consider a more flexible definition and model for propensity score. Imai & Van Dyk (2004) and Hirano & Imbens (2004) propose generalized propensity score methods that allow for the inclusion of the

information provided by covariates to control for selection bias and confounding when treatment assignment is non-binary (Zhao et al. 2012). In support of these generalized propensity scores, Imai & Van Dyk (2004) derive "large-sample" theoretical results of balancing properties and ignorable treatment assignment resembling those of the standard propensity score proposed by Rosenbaum & Rubin (1983). Thus, when treatment assignment is defined as time to treatment, we may use the linear predictor of the Cox proportional hazards model $h(t) = h_0(t) \exp\left(\mathbf{D}^T \boldsymbol{\beta}\right)$ as the generalized propensity score.

The estimated propensity scores are included in the model for the outcome ($\mathbf{Y}$), the model for missing outcome ($\mathbf{M}$), and the model determining principal strata ($\mathbf{G}$), to control for issues of selection bias and confounding. It is noteworthy that inclusion of the propensity scores in the principal strata model is necessary in the absence of randomization, as otherwise the principal effect is likely to be biased. To allow flexibility in the control for selection bias when including each of the propensity scores a linear predictor, the use of quadratic and cubic polynomial higher order terms are also considered when incorporateing propensity score in the models.

4.2.3 Principal Stratification

Even if all outcomes are observed for living patients, the outcome Y_i cannot be measured for those patients who are not alive at the time of outcome measurement, $S_i = 0$. Those outcomes that are not measured due to patient death are considered not defined on the set of real positive numbers, $\mathbf{R}^+$. Following the notation of Zhang & Rubin (2003), we can instead consider the non-observed outcomes to be *, extending our sample space to $\{\mathbf{R}^+, *\}$. In the presence of the censoring of the outcome by death, principal stratification

using post treatment survival status allows for estimation of the treatment effect. Specifically, the Survivor Average Causal Effect (SACE) is defined as the mean difference in the outcomes of treated individuals compared to untreated individuals in the LL stratum, $E(Y_{LL,i}(1)) - E(Y_{LL,i}(0))$. The four potential principal strata are constructed by pairing indicators of survival by treatment scenarios: LL, LD, DL, and DD. Individuals who are in the LL stratum are those who would be alive at time t^o regardless of treatment. Those individuals who are in the LD stratum would be alive if they received the PEG tube but would not be alive if they did not and those who are in the DL stratum would not be alive if they receive PEG treatment, but would be alive if they did not. Finally, individuals in the DD are those who would not be alive at time t^o regardless of treatment.

The probabilities of the four strata (π_{LL}, π_{LD}, π_{DL}, and π_{DD}) can be modeled with a multinomial logit model using a subset of the observed covariate matrix $\mathbf{X}$, $\mathbf{X_2}$, which must include **PS** and may include patient characteristics **D**. The probability of an individual being in principal strata g is given in equation (4.1). As in any multinomial logit model, one category must be selected as a reference group.

$$\pi_{g,i} = P(G_i = g) = \frac{\exp\left(X_{2i}^T \boldsymbol{\alpha}_g\right)}{\sum_{g'} \exp\left(X_{2i}^T \boldsymbol{\alpha}_{g'}\right)} \tag{4.1}$$

However, for any given individual i, we only observe the survival outcome given the observed treatment status. We may define four observed groups based on indicators of treatment and survival at time t^o $O(Z, S)$: $O(1,1)$, $O(1,0)$, $O(0,1)$, $O(0,0)$. These observed groups are composed of mixtures of the principal strata. In other words, $O(1,1)$ is comprised of a mixture of individuals from the LL and LD strata, $O(1,0)$ is comprised of individuals from

the LL and DL strata, $O(0,1)$ is comprised of individuals from the DD and DL strata, and $O(0,0)$ is comprised of individuals from the DL and DD strata.

4.2.4 Missing Outcomes

Thus far, our framework assumes that all those individuals surviving until t^o will have an observed outcome. In studies such as clinical trials or other prospective study designs, it may be appropriate to expect outcome or covariate measurement at planned time-points. However, the reality of retrospective studies, such as those based on disease registries with no interference in patient clinic visits, is that often measurements of patient status are available at sporadic times, and do not always align with the research question of interest. Straightforward options for defining the outcome of interest may either be too broad, by including any outcome observed from baseline until the time of interest, or may be too exclusive, by only including those outcomes observed in a narrow window around the time of interest. This latter definition of the outcome is specific and does allow for a precise analysis of the effect of treatment on outcome, however by excluding those individuals who do not have an outcome in the time period of interest, we squander the information that these individuals may provide.

Instead, we may define the outcome for those individuals who do not have a measurement during the timeframe of interest as missing. This allows for two potential approaches for including the information for the individuals with unobserved outcomes. First, those individuals who do not have an observed outcome but for whom we have information about survival and treatment may contribute to the other parts of the model framework. Namely, these individuals would contribute only to the propensity score and principal stratification

models.

Second, we may introduce a model for the indicator of missing outcome within the principal stratification framework. Explicitly, we model $p(M = 1|G = g, Z = z)$ or $\phi_{g,z}$, the probability of missing outcome using a logistic regression model, in a similar approach to that of Frumento et al. (2012). However, our data introduce some unique challenges that were not of concern for Frumento et al. (2012), such as non-randomized treatment assignment and non-fixed times of outcome measurement. Though the Job Corps data studied by Frumento et al. did present some additional challenges due to issues of non-compliance, the authors were able to treat compliance as a second post-treatment variable and were able to combine this information with employment status to create a single set of six principal strata. Also, because employment status cannot be determined without individual study participation, this post-treatment stratifying variable may be missing in tandem with missing outcome, whereas the retrospective nature of the Emory ALS data ensured that the post-treatment stratifying variable of survival is always observed.

The two modeling approaches presented in this analysis differ by the assumptions placed on the missingness mechanism. In Approach 1, we assume that the missing mechanism is ignorable. For Approach 2, we assume latent ignorability, but since the latent strata are not completely known, the missing mechanism is not ignorable. Thus it is necessary to model the probability of missing outcome data and include it in the framework. We assume this probability differs not only by principal strata but also by treatment status. However, because those individuals who are not alive at the time of outcome measurement cannot have an observed outcome, it is only valid to model the probability of missingness for those individuals in the LL stratum with treat-

ment, the LL stratum without treatment, the LD stratum with treatment, and the DL stratum without treatment. These probabilities are represented in the model below.

$$\phi_{g,z} = Pr(M = 1|G = g, Z = z, S = 1) = \frac{e^{X_3\theta_{g,z}}}{1 + e^{X_3\theta_{g,z}}} \tag{4.2}$$

Regardless of whether a model is introduced for $\phi_{g,z}$, we must adjust the observed groups to reflect the missingness of outcomes. Instead of the four observed groups that are described in Section 4.2.3, we now have six observed groups that are defined as the combinations of $O(Z, S, M)$. Namely the groups include O(1,1,1), O(1,1,0), O(1,0,-), O(0,1,1), O(0,1,0), O(0,0,-). The two original observed groups with individuals surviving until t^o, O(1,1) and O(0,1), are each divided into two further groups based upon whether the outcome is observed or missing. For those individuals who do not survive until the time of outcome measurement, $S = 0$, we need not stratify these groups further by missing indicator.

4.2.5 Bayesian Framework for Estimation and Inference

The outcome Y_i, which is only observed when $S_i = 1$ and $M_i = 0$, is assumed to have a normal distribution, $f_{g,i}$, within each of the principal strata and with parameters and covariates that differ by strata. Specifically, the outcome distributions are defined as $Y_{g,i} \sim N(\mathbf{X}_{1,g}\boldsymbol{\eta}_g, \sigma_g^2)$ for $g \in LL, LD, DL$. $\mathbf{X}_{1,LL}$ includes the column for intercept, one or both of the treatment variables depending on the treatment assignment model considered, and the estimated propensity score corresponding to the treatment assignment model. The outcome models for the LD and DL strata do not include any treatment covariates as the individuals with an observed outcome in each of these strata are either

Table 4.1: Individual observed likelihood by observed groups if G_i is known

	G_i			
Observed Group	LL	LD	DL	DD
$O(1,1,0)$	$(1-\phi_{LL,1,i})\times \pi_{LL,i} f_{LL,i}$	$(1-\phi_{LD,1,i})\times \pi_{LD,i} f_{LD,i}$	-	-
$O(1,1,1)$	$\phi_{LL,1,i}\pi_{LL,i}$	$\phi_{LD,1,i}\pi_{LD,i}$	-	-
$O(1,0,-)$	-	-	$\pi_{DL,i}$	$\pi_{DD,i}$
$O(0,1,0)$	$(1-\phi_{LL,0,i})\times \pi_{LL,i} f_{LL,i}$	-	$(1-\phi_{DL,0,i})\times \pi_{DL,i} f_{DL,i}$	-
$O(0,1,1)$	$\phi_{LL,0,i}\pi_{LL,i}$	-	$\phi_{DL,0,i}\pi_{DL,i}$	-
$O(0,0,-)$	-	$\pi_{LD,i}$	-	$\pi_{DD,i}$

all treated or all untreated respectively. Therefore, $X_{1,LD}$ and $X_{1,DL}$ include columns for intercept and propensity score only.

Using the stratified outcome distributions, the probability of each principal stratum, and the probability missing outcome the structure of the observed data likelihood for any individual and for all possible combinations of Z_i, S_i, and M_i is given in Table 4.1. Each cell value is the likelihood of the observed data if the values of the individual strata are known for the above described modeling approach 2 in which a model for the missing mechanism is introduced into the framework. If instead we employ the first approach, in which latent ignorability is assumed outright, we can drop all $\phi_{g,z}$ from the table.

The conditional probability of $G_i = g$ given the observed data is the ratio of each cell to the total of that row. Rows $O(1,0,-)$ and $O(0,0,-)$ are included in this table for a comprehensive understanding of the possible combinations of treatment and survival, but individuals who fall into these groups do not have outcome data that will contribute to the observed data likelihood since $S_i = 0$. Therefore, individuals in this group will only contribute to the model

for the probability of principal strata, which is reflected in the observed data likelihood in Appendix C.1.

Prior distributions for the specified parameters in the observed data likelihood should be chosen carefully, with thought to distributions that may be informative, proper, and conjugate where appropriate. For this analysis, conjugate multivariate normal and inverse-gamma distributions are assigned for the prior distributions of the different forms of $\boldsymbol{\eta}_g$ and σ_g^2 respectively. The prior distributions for $\boldsymbol{\alpha}_g$ and $\boldsymbol{\theta}_{g,z}$ are non-informative and are proportional to 1. Details of each prior distribution are provided in Appendix C.2.

The posterior distribution of the parameters, the product of the observed data likelihood and the prior distributions, is used for inference on the parameters of interest. Though the principal stratum of each individual is unknown, the observed treatment and survival groups may be used to inform imputation of the principal strata assignments. This may be accomplished via the Data Augmentation (DA) algorithm (Tanner & Wong 1987), in which information about the latent groups (in this case G_i) is imputed and subsequently the posterior parameters distributions are simulated to inference.

The DA algorithm is employed by using two iterative and alternating steps to simulate a complete data likelihood and allow for posterior inference. The first step, the Imputation or I-step, imputes the value of the principal strata G_i for each individual. This is accomplished by using the parameter values $\boldsymbol{\alpha}_g^{(k)}$, $\boldsymbol{\eta}_g^{(k)}$, and $\sigma_g^{2(k)}$, and $\boldsymbol{\theta}_{g,z}^{(k)}$ (for approach 2 only) from the current approximation of posterior (from the kth iteration) to generate $G_i^{(k+1)}$ by using the conditional probabilities that are given by taking the ratio of cell value to row total in Table 4.1. This conditional probabilities, $\rho_{O,i}$ are used in a Bernoulli distribution that imputes individual membership to one of the two principal strata that correspond with the observed group O (see Appendix C.3). More

specifically, at the $(k+1)$ iteration, each individual has a probability of being in a stratum that depends on their observed values $(Z_i,\ S_i,\ M_i,\ Y_i,\ PS_i)$.

The P-step, or Posterior step, is then employed by using the imputed complete data set, and the parameters $\{\boldsymbol{\theta}^{(k)}, \left(\pi_g^{(k)}, \boldsymbol{\eta}_g^{(k)}, \sigma_g^{2(k)}\right)\}$ can be updated to $\{\boldsymbol{\theta}^{(k+1)}, \left(\pi_g^{(k+1)}, \boldsymbol{\eta}_g^{(k+1)}, \sigma_g^{2(k+1)}\right)\}$ by sampling from the full conditional distributions of each parameter. Either the Gibbs Sampler or the Metropolis-Hastings (MH) Algorithm may be employed for sampling, with choice of algorithm influenced by the type of full conditional distribution. The full conditional distributions of each parameter (given the imputed G at each iteration k) are provided in Appendix C.4.

4.3 Application to Emory ALS Clinic Data

The ALS registry dataset includes data from 729 patients who visited the Emory ALS clinic at least once from January 1, 1997 and July 31, 2011. Patients were excluded from the analysis for not having any follow up clinic visits from baseline to outcome measurement or for having extreme survival times (>5 years post-baseline). Characteristics measured at baseline for each individual include sex, site of ALS onset, age at diagnosis, BMI at baseline, and days from diagnosis to first clinic visit. Additionally, some individual characteristics are measured at each clinic visit including forced vital capacity (FVC), change in FVC from baseline, and total number of clinic visits. Those characteristics that are measured as continuous variables (namely age at diagnosis, BMI at baseline, FVC, and change in FVC from baseline) are normalized before inclusion as covariates for the propensity score model, principal strata model, or outcome model. The outcome of interest, is BMI at 18 months post baseline, with a measurement period of 2 months to either side

of this time point.

A comparison of those who receive treatment within one year of follow-up and those who do not among the remaining 623 individuals in the ALS registry is available in Table 4.2. Of the 267 treated patients, 23.6% or 63 individuals are alive 18 months from baseline, while of the 356 untreated individuals, 19.4% or 69 individuals are alive at this time-point (p = 0.240). Baseline measurements of BMI and FVC are not significantly different among the treated and untreated populations were not significant at a level of $\alpha = 0.05$, however, the change in BMI and the change in FVC measurements from baseline to t^o (ΔBMI_{t^o} and ΔFVC_{t^o} respectively) are significantly different among treatment groups with p-values of 0.04 less than 0.001 respectively. In both cases, treated individuals have a greater mean loss in BMI and FVC measurement than untreated individuals over the 18 month time span. NIF score is significantly higher for treated individuals than untreated individuals ($p < 0.001$). Though age at diagnosis seems relatively similar for treated and untreated individuals, there is a significantly higher proportion of females in the treated population ($p = 0.01$) and a significantly lower proportion of spinal onset ALS

Table 4.2: Comparison of demographic and clinical characteristics amongst PEG treated and untreated populations (N=623)

	Treated (N=267)		***Untreated (N=356)***		
	Mean or Percent	*SD or n*	*Mean or Percent*	*SD or n*	*p-value*
Baseline BMI	25.07	5.62	25.62	6.12	0.286
ΔBMI_{t^o}	-1.89	2.57	-1.30	3.54	0.039
Baseline FVC	69.67	26.38	70.61	26.20	0.672
ΔFVC_{t^o}	-28.66	26.17	-15.77	22.18	<0.001
NIF	-28.26	18.67	-44.67	17.07	<0.001
Age at diagnosis	62.91	10.91	62.76	12.17	0.870
Female	50.9%	136	40.4%	144	0.012
Spinal site of onset	47.6%	127	80.9%	288	<0.001
$\Delta T_{DX} > 6$ months	23.2%	62	20.5%	73	0.474
Survival past t^o	23.6%	63	19.4%	69	0.240

Table 4.3: SACE of PEG treatment (with 95% credible intervals) on BMI measured 18 months post-baseline (N=623)

	PEG Treatment Effect Estimate		
	MCAR Method (From Chapter 2)	Approach 1	Approach 2
Time of Treatment	0.13 *(-0.07, 0.34)*	0.18 *(-0.10, 0.42)*	0.18 *(-0.08, 0.46)*
Binary Indicator	1.75 *(-0.32, 3.87)*	4.22 *(0.79, 7.37)*	4.36 *(1.30, 7.86)*

cases ($p < 0.001$). This decrease in the proportion of spinal onset patients along with the lower mean FVC over time and greater proportion of females indicates increased risk of advanced disease in the treated patient population.

4.3.1 Estimation of SACE of PEG Treatment

The Survivor Average Causal Effect (SACE) of PEG treatment is estimated for several modeling scenarios. In addition to the two approaches described in Section 4.2.4, a method that includes all individuals with an outcome measurement prior to t^o is also considered. This method, described in detail as the proposed contributions of Chapter 2, is the equivalent of assuming that the missing observations are missing completely at random (MCAR) and without any regard to the latent strata, which is a fairly strong assumption to make. Also, methods without propensity score adjustment as well as with a monotonicity assumption (removal of the DL stratum) are reviewed. For all analyses, the MCMC algorithm was run for a total of 10,000 iterations, with a burn-in period of 5000 iterations.

Table 4.3 presents a comparison of the different modeling approaches for this analysis. Both modeling Approach 1, which assumes ignorability of the missing mechanism, and Approach 2, in which a model for the missing mechanism is introduced, indicate there is a positive and significant effect of the

Table 4.4: SACE of PEG treatment (with 95% credible intervals) on BMI 18 months post-baseline with and without propensity scores (N=623)

	PEG Treatment Effect Estimate	
	No propensity scores	Approach 2
Time of Treatment	0.30 *(0.02, 0.56)*	0.18 *(-0.08, 0.46)*
Binary Indicator	2.53 *(-0.85, 6.03)*	4.36 *(1.30, 7.86)*

binary indicator of PEG surgery on the outcome of BMI at 18 months post baseline. Holding all else constant and assuming PEG insertion at or shortly after baseline, the BMI at 18 months increases by about 4 units for those individuals who have treatment when compared to those who are not treated. The effect of time to treatment is also positive, indicating there may be some advantage to having PEG insertion at later time points, but this effect estimate is not significant for either approach. The naïve analysis does not find a significant effect of time to treatment or binary indicator of treatment on BMI at 18 months.

If we do not control for selection bias or confounding by means of generalized propensity scores, we find that the results are slightly different. Table 4.4 includes the results of Approach 2 without propensity score adjustment and compares them to those with quadratic propensity score terms included in the model. While the direction of the effect estimates both models remain the same, there are some interesting differences. The effect estimate of the binary indicator of treatment is smaller in magnitude and no longer significant when no propensity scores are included in the model. However, the effect estimate of time to treatment actually increases and is significant when propensity scores are not used.

Finally, though there are some slight changes in effect magnitude when employing the monotonicity assumption (Table 4.3), the directions and sig-

Table 4.5: SACE of PEG treatment (with 95% credible intervals) on BMI measured 18 months post-baseline when a monotonicity assumption is employed (N=623)

	PEG Treatment Effect Estimate		
	MCAR Method (From Chapter 2)	Approach 1	Approach 2
Time of Treatment	0.14 *(-0.08, 0.35)*	0.20 *(-0.05, 0.43)*	0.20 *(-0.01, 0.42)*
Binary Indicator	1.76 *(-0.44, 3.87)*	3.29 *(1.02, 5.62)*	3.28 *(1.21, 5.74)*

nificance of the effect estimates remain mostly the same. The most notable difference is that in the method assuming MCAR, both the effect of PEG as a binary indicator and time to treatment are now not significant. The results of modeling approaches 1 and 2 are largely the same as without the monotonicity assumption.

Though the results in this model with time from treatment and binary indicator of treatment seem promising in providing a positive treatment effect of PEG, a careful interpretation of the treatment effect estimates is necessary. While there is a positive effect of the binary indicator, this increase in BMI is for those individuals who have PEG insertion at baseline ($T_Z = 0$). The effect of time to treatment is additive to the binary treatment indicator at any time, so the results indicate that the later individuals have surgery, the more beneficial it may be. However, in practice there seems to be an ideal time, prior to extensive disease progression, before which PEG insertion is beneficial. After the disease reaches a certain point of morbidity, physicians note that the treatment offers little to no (and even negative) benefit.

4.4 Discussion

Making adjustments for missing outcome data within the context of causal inference frameworks require strong assumptions about the ignorability of the missing mechanism and creativity in the modeling framework. In this analysis, the similar results of approaches 1 and 2 indicate that latent ignorability is likely a suitable assumption for this data, though we cannot fully test this assumption. When in doubt, the further stratification by missingness of outcome data and within the principal stratification framework assures flexibility in the assumptions imposed on the missing mechanism.

Additionally, the use of propensity scores within the principal stratification framework, whether simple or complex, allows for the estimation of an unbiased principal effect of treatment for observational data. This may also be true for randomized data in which the assumption of no unmeasured confounding is suspect. However, the reliability removing bias via propensity scores relies on the assumption of strongly ignorable treatment assignment must hold, which means there must be no unmeasured confounders. One important confounder that was not available for this analysis is the Revised ALS Functional Rating Scale (ALSFRS-R) score, which is a validated instrument for measuring the progression of ALS. This confounder could be especially helpful in achieveing ignorability when conditioning on propensity scores. Future applications of the proposed methods to other data with richer measurements of confounders should further demonstrate the reduction of selection bias and confounding.

Overall, the results presented from the application to the ALS data are not sensitive to the assumption of monotonicity. In the data application, this may be due to the small proportion of individuals in DL strata when all four strata are considered. When monotonicity is not assumed, most patients are in the LL and DD strata, with a LD and DL strata comprising less than

20% of the individuals in total. It is conceivable then that reallocating such a small proportion of individuals when removing the DL stratum would likely not substantially change the effect estimates of the other strata.

Propensity scores are included as linear predictors of the outcome and principal strata models, which can be quite restrictive when controlling for selection bias and confounding. Adding higher order terms does add some flexibility the parametric model, but other non-polynomial basis functions may be considered in the future for more effective control for bias. Other options that would remove the parametric restrictions of model adjustment include matching by propensity scores or inverse probability weighting.

Future consideration may also be given to jointly modeling the propensity score with the outcome model, missing data model, and principal strata model in the Bayesian framework. This would allow the quantities observed by in each of these three models to affect the posterior of propensity score in each MCMC iteration. While this could provide a more robust propensity score adjustment, Zigler et al. (2013) show that the feedback between model stages in joint modeling can cause biased causal effect estimates if individual covariates are not also adjusted for in the outcome model. This bias should be accounted for if joint modeling of the four models of outcome, missing mechanism, propensity scores, and principal strata is proposed.

Chapter 5

Future Directions for Research

The rich complications of data of the ALS registry could inspire a multitude of analytical projects. One logical extension of the methods described in Chapters 2 and 4 would include considering time-dependent covariates and outcome in the analysis. This would require careful consideration of a time-varying propensity score, perhaps using similar time-varying coefficient Cox Proportional Hazards modeling as in Chapter 3, to control for selection bias over the course of the follow-up period. Also, rather than including propensity score as a model covarariate, methods using propensity score matching would be explored. Matching, however, would require careful thought to the model and the implications of selected controls.

Consideration could be also be given to novel matching algorithms or additional methods utilizing propensity score matched pairs may be considered as extensions of Chapter 3. Matching algorithms that increase effiency and decrease time could be beneficial to the scientific community. Also, other methods of profile matching of individuals over time could be allow a comparison of individuals that takes into account progression of disease. Random effect models, with strata defined by matched pairs, could be introduced for post matching analysis of treatment effect.

Appendix A

Appendix for Chapter 2

A.1 Observed Data Likelihood

$$
\begin{aligned}
P(Y|S,Z,G,D,PS) \propto & \prod_{i\in O(1,1)} \left\{ \frac{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}}\sigma_{LL}^{-1} e^{\frac{\left(Y_i-\mathbf{X}_{1,LL,i}\boldsymbol{\eta}_{LL}\right)^2}{2\sigma_{LL}^2}} + \sigma_{LD}^{-1} e^{\frac{\left(Y_i-\mathbf{X}_{1,LD,i}\boldsymbol{\eta}_{LD}\right)^2}{2\sigma_{LD}^2}}}{1+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}}} \right\} \\
& \times \prod_{i\in O(0,1)} \left\{ \frac{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}}\sigma_{LL}^{-1} e^{\frac{\left(Y_i-\mathbf{X}_{1,LL,i}\boldsymbol{\eta}_{LL}\right)^2}{2\sigma_{LL}^2}} + e^{\mathbf{X}_{2i}\boldsymbol{\alpha}_{DL}}\sigma_{DL}^{-1} e^{\frac{\left(Y_i-\mathbf{X}_{1,DL,i}\boldsymbol{\eta}_{DL}\right)^2}{2\sigma_{DL}^2}}}{1+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}}} \right\} \\
& \times \prod_{i\in O(1,0)} \left\{ \frac{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}}}{1+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}}} \right\} \\
& \times \prod_{i\in O(0,0)} \left\{ \frac{1+\mathbf{X}_{2,i}e^{\boldsymbol{\alpha}_{DD}}}{1+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}}} \right\}
\end{aligned}
$$

A.2 Prior Distributions of Parameters

1. $p\left(\boldsymbol{\alpha}_g\right) \propto 1$

2. $\boldsymbol{\eta}_g \sim Normal_p\left(\boldsymbol{\mu}_g, \sigma_g^2\mathbf{V}_g\right)$
   $p\left(\boldsymbol{\eta}_g\right) \propto |\sigma_g^2 V_g|^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_g^2}\left(\boldsymbol{\eta}_g-\boldsymbol{\mu}_g\right)^T \mathbf{V}_g^{-1}\left(\boldsymbol{\eta}_g-\boldsymbol{\mu}_g\right)}$

3. $\sigma_g^2 \sim InverseGamma\left(\nu_g, \omega_g\right)$
   $p\left(\sigma_g^2\right) \propto \sigma_g^{2(-\nu_{LL}-1)} exp\left(-\frac{\omega_{LL}}{\sigma_{LL}^2}\right)$

A.3 Imputation Probabilities for I-Step

1. For $O(1,1)$:

$$
\begin{aligned}
P\left(G_i^{(k+1)} = LL\right) = \rho_{11,i}^{(k+1)} &= \frac{\pi_{LL,i}^{(k)} f_{LL,i}^{(k)}}{\pi_{LL,i}^{(k)} f_{LL,i}^{(k)} + \pi_{LD,i}^{(k)} f_{LD,i}^{(k)}} \\
&= \frac{\frac{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}}}{\sigma_{LL}^{(k)}} \exp\left(\frac{(Y_i-\mathbf{X}_{1,LL,i}\boldsymbol{\eta}_{LL})^2}{2\sigma_{LL}^2}\right)}{\frac{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}}}{\sigma_{LL}^{(k)}} \exp\left(\frac{(Y_i-\mathbf{X}_{1,LL,i}\boldsymbol{\eta}_{LL})^2}{2\sigma_{LL}^2}\right) + \frac{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LD}}}{\sigma_{LD}^{(k)}} \exp\left(\frac{(Y_i-\mathbf{X}_{1,LD,i}\boldsymbol{\eta}_{LD})^2}{2\sigma_{LD}^2}\right)} \\
P\left(G_i^{(k+1)} = LD\right) &= 1 - \rho_{11,i}^{(k+1)}
\end{aligned}
$$

2. For $O(0,1)$:

$$P\left(G_i^{(k+1)} = LL\right) = \rho_{01,i}^{(k+1)} = \frac{\pi_{LL,i}^{(k)} f_{LL,i}^{(k)}}{\pi_{LL,i}^{(k)} f_{LL,i}^{(k)} + \pi_{DL,i}^{(k)} f_{DL,i}^{(k)}}$$
$$= \frac{\frac{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}}}{\sigma_{LL}^{(k)}} \exp\left(\frac{(Y_i - \mathbf{X}_{1,LL,i}\boldsymbol{\eta}_{LL})^2}{2\sigma_{LL}^2}\right)}{\frac{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}}}{\sigma_{LL}^{(k)}} \exp\left(\frac{(Y_i - \mathbf{X}_{1,LL,i}\boldsymbol{\eta}_{LL})^2}{2\sigma_{LL}^2}\right) + \frac{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}}}{\sigma_{DL}^{(k)}} \exp\left(\frac{(Y_i - \mathbf{X}_{1,DL,i}\boldsymbol{\eta}_{DL})^2}{2\sigma_{DL}^2}\right)}$$
$$P\left(G_i^{(k+1)} = DL\right) = 1 - \rho_{01,i}^{(k+1)}$$

3. For $O(1,0)$:

$$P\left(G_i^{(k+1)} = DD\right) = \rho_{10,i}^{(k+1)} = \frac{\pi_{DD,i}^{(k)}}{\pi_{DD,i}^{(k)} + \pi_{DL,i}^{(k)}}$$
$$= \frac{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}}}{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}} + e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}}}$$
$$P\left(G_i^{(k+1)} = DL\right) = 1 - \rho_{10,i}^{(k+1)}$$

4. For $O(0,0)$:

$$P\left(G_i^{(k+1)} = DD\right) = \rho_{00,i}^{(k+1)} = \frac{\pi_{DD,i}^{(k)}}{\pi_{DD,i}^{(k)} + \pi_{LD,i}^{(k)}}$$
$$= \frac{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}}}{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}} + e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LD}}}$$
$$P\left(G_i^{(k+1)} = LD\right) = 1 - \rho_{00,i}^{(k+1)}$$

A.4 Full Conditional Distributions of Parameters of Interest in the P-step of the Data Augmentation Algorithm

A.4.1 Full Conditional Distributions of $\boldsymbol{\alpha}_g$

$$P\left(\boldsymbol{\alpha}_{LL}|\mathbf{X}_2,\mathbf{G}^{(k+1)},\boldsymbol{\alpha}_{DL}^{(k)},\boldsymbol{\alpha}_{DD}^{(k)}\right)\propto\prod_{G_i^{(k+1)}=LL}\frac{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}}}{1+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}^{(k)}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}^{(k)}}}$$
$$\times\prod_{G_i^{(k+1)}=LD}\frac{1}{1+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}^{(k)}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}^{(k)}}}$$
$$\times\prod_{G_i^{(k+1)}=DL}\frac{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}^{(k)}}}{1+e^{X_{2,i}\boldsymbol{\alpha}_{LL}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}^{(k)}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}^{(k)}}}$$
$$\times\prod_{G_i^{(k+1)}=DD}\frac{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}^{(k)}}}{1+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}^{(k)}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}^{(k)}}}$$

$$P\left(\boldsymbol{\alpha}_{DL}|\mathbf{X}_2,\mathbf{G}^{(k+1)},\boldsymbol{\alpha}_{LL}^{(k+1)},\boldsymbol{\alpha}_{DD}^{(k)}\right)\propto\prod_{G_i^{(k+1)}=LL}\frac{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}^{(k+1)}}}{1+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}^{(k+1)}}+e^{X_{2,i}\boldsymbol{\alpha}_{DL}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}^{(k)}}}$$
$$\times\prod_{G_i^{(k+1)}=DL}\frac{1}{1+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}^{(k+1)}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}^{(k)}}}$$
$$\times\prod_{G_i^{(k+1)}=DL}\frac{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}}}{1+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}^{(k+1)}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}^{(k)}}}$$
$$\times\prod_{G_i^{(k+1)}=DD}\frac{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}^{(k)}}}{1+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}^{(k+1)}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}^{(k)}}}$$

$$P\left(\boldsymbol{\alpha}_{DD}|\mathbf{X}_2,\mathbf{G}^{(k+1)},\boldsymbol{\alpha}_{LL}^{(k+1)},\boldsymbol{\alpha}_{DL}^{(k+1)}\right) \propto \prod_{G_i^{(k+1)}=LL} \frac{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}^{(k+1)}}}{1+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}^{(k+1)}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}^{(k+1)}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}}}$$
$$\times \prod_{G_i^{(k+1)}=LD} \frac{1}{1+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}^{(k+1)}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}^{(k+1)}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}}}$$
$$\times \prod_{G_i^{(k+1)}=DL} \frac{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}^{(k+1)}}}{1+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}^{(k+1)}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}^{(k+1)}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}}}$$
$$\times \prod_{G_i^{(k+1)}=DD} \frac{e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}}}{1+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{LL}^{(k+1)}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DL}^{(k+1)}}+e^{\mathbf{X}_{2,i}\boldsymbol{\alpha}_{DD}}}$$

A.4.2 Full Conditional Distributions of η_g

$$\boldsymbol{\eta}_g|\mathbf{Y},\mathbf{X}_{1,g},\mathbf{S},\mathbf{Z},G^{(k+1)},\sigma_g^{2(k)} \sim N_p\left(\boldsymbol{\mu}_g^*,\sigma_g^{2(k)}\left(\mathbf{X}_{1,g}^T\mathbf{X}_{1,g}+\mathbf{V}_g^{-1}\right)^{-1}\right)$$
$$\boldsymbol{\mu}_g^* = \left(\mathbf{X}_{1,g}^T\mathbf{X}_{1,g}+\mathbf{V}_g^{-1}\right)^{-1}\left(\mathbf{X}_{1,g}^T\mathbf{Y}+\mathbf{V}_g^{-1}\boldsymbol{\mu}_g\right)$$

A.4.3 Full Conditional Distributions of σ_g^2

$$\sigma_g^2|\mathbf{Y},\mathbf{X}_{1,g},\mathbf{S},\mathbf{Z},G^{(k+1)},\boldsymbol{\eta}_g^{(k+1)} \sim IG\left(\nu_g+\frac{n_g}{2}+\frac{1}{2},\omega_g^*\right)$$
$$\omega_g^* = \omega_g+\frac{1}{2}\left[\left(\boldsymbol{\eta}_g^{(k+1)}-\boldsymbol{\mu}_g\right)^T\mathbf{V}_g^{-1}\left(\boldsymbol{\eta}_g^{(k+1)}-\mu_{LL}\right)+\left(\mathbf{Y}-\mathbf{X}_{1,g}\boldsymbol{\eta}_g^{(k+1)}\right)^T\left(\mathbf{Y}-\mathbf{X}_{1,g}\boldsymbol{\eta}_g^{(k+1)}\right)\right]$$

A.5 Summary of Patient Characteristics at Several Time-points

Table A.1: Comparison of PEG among treated and untreated populations times post-baseline

	3 Months			**6 Months**		
	Treated (N=80)	*Untreated (N=404)*	*p-value*	*Treated (N=130)*	*Untreated (N=408)*	*p-value*
BMI	23.2 (6.7)	26 (5.7)	< 0.01	23.6 (5.8)	25.8 (5.6)	< 0.01
Baseline (BL) BMI	23.4 (6.9)	26 (5.8)	< 0.01	24.2 (6.2)	25.9 (5.7)	0.01
ΔBMI_{t^o}	-0.3 (0.8)	-0.1 (0.6)	0.02	-0.7 (1.1)	-0.1 (2.3)	< 0.01
FVC	50.8 (23.4)	73.4 (24)	< 0.01	51 (22.4)	71.1 (24.7)	< 0.01
Baseline (BL) FVC	53.3 (22.7)	76.7 (24.8)	< 0.01	59.2 (25.6)	76.4 (25.1)	< 0.01
ΔFVC_{t^o}	-3.5 (7.8)	-3.5 (10.9)	0.94	-9.6 (15.1)	-5.6 (14.1)	0.01
Age at Diagnosis	67 (12)	61.9 (11.1)	< 0.01	66.4 (10.4)	61.8 (11.3)	< 0.01
$\Delta T_{DX} > 30$ days	0.09 (7)	0.11 (44)	0.71	0.13 (17)	0.15 (62)	0.65
Prop. surviving at t*	0.75 (60)	0.90 (364)	< 0.01	0.61 (80)	0.77 (318)	< 0.01
Prop. of Females	0.63 (50)	0.41 (164)	< 0.01	0.55 (72)	0.41 (168)	0.01
Prop. of Spinal Onset	0.35 (28)	0.72 (292)	< 0.01	0.36 (48)	0.75 (307)	< 0.01
	1 Year			**18 Months**		
	Treated (N=200)	*Untreated (N=380)*	*p-value*	*Treated (N=242)*	*Untreated (N=361)*	*p-value*
BMI	23.9 (5.5)	24.9 (5.6)	0.05	23.7 (5.2)	24.5 (5.9)	0.11
Baseline (BL) BMI	24.9 (5.9)	25.6 (5.7)	0.19	25 (5.7)	25.6 (5.9)	0.24
ΔBMI_{t^o}	-1.1 (2.1)	-0.7 (2.9)	0.04	-1.6 (2.3)	-1 (3.4)	0.02
FVC	46.8 (21.4)	63.5 (25.1)	< 0.01	44.3 (20.3)	58.5 (25.3)	< 0.01
Baseline (BL) FVC	65.2 (25.6)	74 (25.8)	< 0.01	68.5 (26.4)	72.1 (26.3)	0.11
ΔFVC_{t^o}	-18.9 (22.1)	-10.4 (17)	< 0.01	-24.9 (24)	-13.2 (21.4)	< 0.01
Age at Diagnosis	64.9 (10.2)	61.9 (12.2)	< 0.01	63.6 (10.7)	62.5 (12.1)	0.24
$\Delta T_{DX} > 30$ days	0.20 (41)	0.18 (71)	0.68	0.22 (54)	0.19 (69)	0.39
Prop. surviving at t*	0.41 (83)	0.54 (206)	< 0.01	0.33 (80)	0.33 (122)	0.92
Prop. of Females	0.51 (103)	0.41 (156)	0.02	0.50 (122)	0.40 (148)	0.03
Prop. of Spinal Onset	0.43 (86)	0.80 (304)	< 0.01	0.45 (111)	0.80 (292)	< 0.01
	2 Years			**3 Years**		
	Treated (N=266)	*Untreated (N=359)*	*p-value*	*Treated (N=290)*	*Untreated (N=364)*	*p-value*
BMI	23.5 (5.2)	24 (6)	0.36	23.4 (5.1)	23.9 (5.9)	0.34
Baseline (BL) BMI	25.1 (5.6)	25.7 (6.1)	0.24	25.3 (5.6)	25.6 (6.1)	0.45
ΔBMI_{t^o}	-1.8 (2.5)	-1.3 (3.8)	0.11	-2.1 (2.7)	-1.3 (3.7)	0.01
FVC	41.8 (19.5)	56 (23.9)	< 0.01	39.7 (20.5)	53.6 (23.7)	< 0.01
Baseline (BL) FVC	69.7 (26.3)	71.2 (26.2)	0.47	70.9 (26.2)	70 (26.5)	0.65
ΔFVC_{t^o}	-28.1 (25.9)	-14.9 (20.8)	¡0.01	-31.3 (27.6)	-16 (21.9)	< 0.01
Age at Diagnosis	63 (10.9)	62.6 (12.3)	0.61	62 (11.2)	62.9 (12.2)	0.32
$\Delta T_{DX} > 30$ days	0.22 (60)	0.20 (73)	0.57	0.24 (72)	0.18 (69)	0.09
Prop. surviving at t*	0.22 (61)	0.19 (71)	0.39	0.10 (30)	0.06 (22)	0.06
Prop. of females	0.51 (137)	0.40 (145)	0.01	0.49 (144)	0.40 (147)	0.02
Prop. of spinal onset	0.47 (126)	0.81 (292)	< 0.01	0.50 (146)	0.80 (293)	< 0.01

Appendix B

Appendix for Chapter 3

B.1 Proof of Propositions 1 and 2

We prove Propositions 1 and 2 based on the treatment assignment model (3.1) and it is straightforward to extend this proof to other types of models. From equation (3.1) and the definition of the Propensity Process $\theta(\cdot)$, we have

$$\begin{aligned} h_{T_Z}(t|\mathbf{X}_1, \mathbf{X}_2(t)) &= h_0(t) exp\left\{\boldsymbol{\beta}_1^T \mathbf{X}_1 + \boldsymbol{\beta}_2^T \mathbf{X}_2(t)\right\} \\ &= h_0(t) exp\left\{\theta(t)\right\}. \end{aligned}$$

From equation (3.2), we have

$$\begin{aligned} f_{T_Z}(t|\mathbf{X}_1, \mathbf{X}_2(t), \theta(t)) &= f_{T_Z}(t|\mathbf{X}_1, \mathbf{X}_2(t)) \\ &= h_{T_Z}(t|\mathbf{X}_1, \mathbf{X}_2(t)) exp\{-\int_0^t h_{T_Z}(t|\mathbf{X}_1, \mathbf{X}_2(t))dt\} \\ &= h_0(t) exp\left\{\theta(t)\right\} exp\left\{-\int_0^t h_0(t) exp\{\theta(t)\}dt\right\} \\ &= f_{T_Z}(t|\theta(t)), \text{ for } t \in [0, t^o), \end{aligned} \tag{B.1}$$

where the first equality is due to the fact that $\theta(t)$ is redundant given $\mathbf{X}_1$ and $\mathbf{X}_2(t)$. The result in Proposition 1 follows immediately, i.e., conditional on $\overline{\theta}$, the distribution of T_Z is independent of $\mathbf{X}_1$ and $\overline{\mathbf{X}}_2$.

In addition, it follows from (B.1) that $f_{T_Z}(t|\mathbf{X}_1, \mathbf{X}_2(t), \theta(t)) = f_{T_Z}(t|\theta(t))$ for $t < t^*$ and

$$\begin{aligned} P(T_Z > t^*|\mathbf{X}_1, \overline{\mathbf{X}}_2(t^*), \overline{\theta}(t^*)) &= 1 - P(T_Z \leq t^*|\mathbf{X}_1, \overline{\mathbf{X}}_2(t^*), \overline{\theta}(t^*)) \\ &= 1 - P(T_Z \leq t^*|\overline{\theta}(t^*)) \\ &= P(T_Z > t^*|\overline{\theta}(t^*)), \end{aligned}$$

where the second equality is due to (B.1). The result in Proposition 2 follows immediately, i.e., conditional on $\overline{\theta}(t^*)$, the distribution of $T_Z(t^*)$ is independent of $\mathbf{X}_1$ and $\overline{\mathbf{X}}_2(t^*)$.

Appendix C

Appendix for Chapter 4

C.1 Observed Data Likelihood for Modelling Approach 2

$$
\begin{aligned}
P(Y|S,M,G,Z,D,PS) \propto & \prod_{i\in O(1,1,0)} \left\{ \frac{\frac{1}{1+e^{X_{3,i}\theta_{LL,1}}} e^{X_{2,i}\alpha_{LL}} \sigma_{LL}^{-1} e^{\frac{\left(Y_i - X_{1,LL,i}\eta_{LL}\right)^2}{2\sigma_{LL}^2}}}{1 + e^{X_{2,,i}\alpha_{LL}} + e^{X_{2,i}\alpha_{DL}} + e^{X_{2,i}\alpha_{DD}}} + \right. \\
& \left. \frac{\frac{1}{1+e^{X_{3,i}\theta_{LD,1}}} \sigma_{LD}^{-1} e^{\frac{\left(Y_i - X_{1,LD,i}\eta_{LD}\right)^2}{2\sigma_{LD}^2}}}{1 + e^{X_{2,i}\alpha_{LL}} + e^{X_{2,i}\alpha_{DL}} + e^{X_{2,i}\alpha_{DD}}} \right\} \\
& \times \prod_{i\in O(0,1,0)} \left\{ \frac{\frac{1}{1+e^{X_{3,i}\theta_{LL,0}}} e^{X_{2,i}\alpha_{LL}} \sigma_{LL}^{-1} e^{\frac{\left(Y_i - X_{1,LL,i}\eta_{LL}\right)^2}{2\sigma_{LL}^2}}}{1 + e^{X_{2,i}\alpha_{LL}} + e^{X_{2,i}\alpha_{DL}} + e^{X_{2,i}\alpha_{DD}}} + \right. \\
& \left. \frac{\frac{1}{1+e^{X_{3,i}\theta_{DL,0}}} e^{X_{2,i}\alpha_{DL}} \sigma_{DL}^{-1} e^{\frac{\left(Y_i - X_{1,DL,i}\eta_{DL}\right)^2}{2\sigma_{DL}^2}}}{1 + e^{X_{2,i}\alpha_{LL}} + e^{X_{2,i}\alpha_{DL}} + e^{X_{2,i}\alpha_{DD}}} \right\} \\
& \times \prod_{i\in O(1,1,1)} \left\{ \frac{\frac{e^{X_{3,i}\theta_{LD,1}}}{1+e^{X_{3,i}\theta_{LD,1}}} + \frac{e^{X_{3,i}\theta_{LL,1}}}{1+e^{X_{3,i}\theta_{LL,1}}} e^{X_{2,i}\alpha_{LL}}}{1 + e^{X_{2,i}\alpha_{LL}} + e^{X_{2,i}\alpha_{DL}} + e^{X_{2,i}\alpha_{DD}}} \right\} \\
& \times \prod_{i\in O(0,1,1)} \left\{ \frac{\frac{e^{X_3\theta_{LL,0}}}{1+e^{X_{3,i}\theta_{LL,0}}} e^{X_{2,i}\alpha_{LL}} + \frac{e^{X_{3,i}\theta_{DL,0}}}{1+e^{X_{3,i}\theta_{DL,0}}} e^{X_{2,i}\alpha_{DL}}}{1 + e^{X_{2,i}\alpha_{LL}} + e^{X_{2,i}\alpha_{DL}} + e^{X_{2,i}\alpha_{DD}}} \right\} \\
& \times \prod_{i\in O(1,0,-)} \left\{ \frac{e^{X_{2i,i}\alpha_{DL}} + e^{X_{2,i}\alpha_{DD}}}{1 + e^{X_{2,i}\alpha_{LL}} + e^{X_{2,i}\alpha_{DL}} + e^{X_{2,i}\alpha_{DD}}} \right\} \\
& \times \prod_{i\in O(0,0,-)} \left\{ \frac{1 + e^{X_{2,i}\alpha_{DD}}}{1 + e^{X_{2,i}\alpha_{LL}} + e^{X_{2,i}\alpha_{DL}} + e^{X_{2,i}\alpha_{DD}}} \right\}
\end{aligned}
$$

C.2 Prior Distributions of Parameters

1. $p(\alpha_g) \propto 1$
2. $p(\theta_{g,z}) \propto 1$
3. $\eta_g \sim Normal_p(\mu_g, \sigma_g^2 V_g)$
   $p(\eta_g) \propto |\sigma_g^2 V_g|^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_g^2}(\eta_g-\mu_g)^T V_g^{-1}(\eta_g-\mu_g)}$
4. $\sigma_g^2 \sim InverseGamma(\nu_g, \omega_g)$
   $p(\sigma_g^2) \propto \sigma_g^{2(-\nu_{LL}-1)} exp\left(-\frac{\omega_{LL}}{\sigma_{LL}^2}\right)$

C.3 Imputation Probabilities for I-Step

1. For $O(1,1,0)$:

$$P\left(G_i^{(k+1)} = LL\right) = \rho_{110,i}^{(k+1)} = \frac{(1-\phi_{LL,1,i}^{(k)})\pi_{LL,i}^{(k)} f_{LL,i}^{(k)}}{(1-\phi_{LL,1,i}^{(k)})\pi_{LL,i}^{(k)} f_{LL,i}^{(k)} + (1-\phi_{LD,1,i}^{(k)})\pi_{LD,i}^{(k)} f_{LD,i}^{(k)}}$$
$$P\left(G_i^{(k+1)} = LD\right) = 1 - \rho_{110,i}^{(k+1)}$$

2. For $O(0,1,0)$:

$$P\left(G_i^{(k+1)} = LL\right) = \rho_{010,i}^{(k+1)} = \frac{(1-\phi_{LL,0,i}^{(k)})\pi_{LL,i}^{(k)} f_{LL,i}^{(k)}}{(1-\phi_{LL,0,i}^{(k)})\pi_{LL,i}^{(k)} f_{LL,i}^{(k)} + (1-\phi_{DL,0,i}^{(k)})\pi_{DL,i}^{(k)} f_{DL,i}^{(k)}}$$
$$P\left(G_i^{(k+1)} = DL\right) = 1 - \rho_{010,i}^{(k+1)}$$

3. For $O(1,1,1)$:

$$P\left(G_i^{(k+1)} = LL\right) = \rho_{111,i}^{(k+1)} = \frac{\phi_{LL,1,i}^{(k)}\pi_{LL,i}^{(k)}}{\phi_{LL,1,i}^{(k)}\pi_{LL,i}^{(k)} + \phi_{LD,1,i}^{(k)}\pi_{LD,i}^{(k)}}$$
$$P\left(G_i^{(k+1)} = DL\right) = 1 - \rho_{111,i}^{(k+1)}$$

4. For $O(0,1,1)$:

$$P\left(G_i^{(k+1)} = LL\right) = \rho_{011,i}^{(k+1)} = \frac{\phi_{LL,0,i}^{(k)}\pi_{LL,i}^{(k)}}{\phi_{LL,0,i}^{(k)}\pi_{LL,i}^{(k)} + \phi_{DL,0,i}^{(k)}\pi_{DL,i}^{(k)}}$$
$$P\left(G_i^{(k+1)} = DL\right) = 1 - \rho_{011,i}^{(k+1)}$$

5. For $O(1,0,-)$:

$$P\left(G_i^{(k+1)} = DD\right) = \rho_{10-,i}^{(k+1)} = \frac{\pi_{DD,i}^{(k)}}{\pi_{DD,i}^{(k)} + \pi_{DL,i}^{(k)}}$$
$$P\left(G_i^{(k+1)} = DL\right) = 1 - \rho_{10-,i}^{(k+1)}$$

6. For $O(0,0,-)$:

$$P\left(G_i^{(k+1)} = DD\right) = \rho_{00-,i}^{(k+1)} = \frac{\pi_{DD,i}^{(k)}}{\pi_{DD,i}^{(k)} + \pi_{LD,i}^{(k)}}$$
$$P\left(G_i^{(k+1)} = LD\right) = 1 - \rho_{00-,i}^{(k+1)}$$

C.4 Full Conditional Distributions of Parameters of Interest in the P-step of the Data Augmentation Algorithm

C.4.1 Full Conditional Distributions of α_g

$$P\left(\alpha_{LL}|X_2, G^{(k+1)}, \alpha_{DL}^{(k)}, \alpha_{DD}^{(k)}\right) \propto \prod_{G_i^{(k+1)}=LL} \frac{e^{X_{2,i}\alpha_{LL}}}{1+e^{X_{2,i}\alpha_{LL}}+e^{X_{2,i}\alpha_{DL}^{(k)}}+e^{X_{2,i}\alpha_{DD}^{(k)}}}$$
$$\times \prod_{G_i^{(k+1)}=LD} \frac{1}{1+e^{X_{2,i}\alpha_{LL}}+e^{X_{2,i}\alpha_{DL}^{(k)}}+e^{X_{2,i}\alpha_{DD}^{(k)}}}$$
$$\times \prod_{G_i^{(k+1)}=DL} \frac{e^{X_{2,i}\alpha_{DL}^{(k)}}}{1+e^{X_{2,i}\alpha_{LL}}+e^{X_{2,i}\alpha_{DL}^{(k)}}+e^{X_{2,i}\alpha_{DD}^{(k)}}}$$
$$\times \prod_{G_i^{(k+1)}=DD} \frac{e^{X_{2,i}\alpha_{DD}^{(k)}}}{1+e^{X_{2,i}\alpha_{LL}}+e^{X_{2,i}\alpha_{DL}^{(k)}}+e^{X_{2,i}\alpha_{DD}^{(k)}}}$$

$$P\left(\alpha_{DL}|X_2, G^{(k+1)}, \alpha_{LL}^{(k+1)}, \alpha_{DD}^{(k)}\right) \propto \prod_{G_i^{(k+1)}=LL} \frac{e^{X_{2,i}\alpha_{LL}^{(k+1)}}}{1+e^{X_{2,i}\alpha_{LL}^{(k+1)}}+e^{X_{2,i}\alpha_{DL}}+e^{X_{2,i}\alpha_{DD}^{(k)}}}$$
$$\times \prod_{G_i^{(k+1)}=LD} \frac{1}{1+e^{X_{2,i}\alpha_{LL}^{(k+1)}}+e^{X_{2,i}\alpha_{DL}}+e^{X_{2,i}\alpha_{DD}^{(k)}}}$$
$$\times \prod_{G_i^{(k+1)}=DL} \frac{e^{X_{2,i}\alpha_{DL}}}{1+e^{X_{2,i}\alpha_{LL}^{(k+1)}}+e^{X_{2,i}\alpha_{DL}}+e^{X_{2,i}\alpha_{DD}^{(k)}}}$$
$$\times \prod_{G_i^{(k+1)}=DD} \frac{e^{X_{2,i}\alpha_{DD}^{(k)}}}{1+e^{X_{2,i}\alpha_{LL}^{(k+1)}}+e^{X_{2,i}\alpha_{DL}}+e^{X_{2,i}\alpha_{DD}^{(k)}}}$$

$$P\left(\alpha_{DD}|X_2, G^{(k+1)}, \alpha_{LL}^{(k+1)}, \alpha_{DL}^{(k+1)}\right) \propto \prod_{G_i^{(k+1)}=LL} \frac{e^{X_{2,i}\alpha_{LL}^{(k+1)}}}{1+e^{X_{2,i}\alpha_{LL}^{(k+1)}}+e^{X_{2,i}\alpha_{DL}^{(k+1)}}+e^{X_{2,i}\alpha_{DD}}}$$
$$\times \prod_{G_i^{(k+1)}=LD} \frac{1}{1+e^{X_{2,i}\alpha_{LL}^{(k+1)}}+e^{X_{2,i}\alpha_{DL}^{(k+1)}}+e^{X_{2,i}\alpha_{DD}}}$$
$$\times \prod_{G_i^{(k+1)}=DL} \frac{e^{X_{2,i}\alpha_{DL}^{(k+1)}}}{1+e^{X_{2,i}\alpha_{LL}^{(k+1)}}+e^{X_{2,i}\alpha_{DL}^{(k+1)}}+e^{X_{2,i}\alpha_{DD}}}$$
$$\times \prod_{G_i^{(k+1)}=DD} \frac{e^{X_{2,i}\alpha_{DD}}}{1+e^{X_{2,i}\alpha_{LL}^{(k+1)}}+e^{X_2\alpha_{DL}^{(k+1)}}+e^{X_{2,i}\alpha_{DD}}}$$

C.4.2 Full Conditional Distributions of $\theta_{g,z}$

$$P\left(\theta_{g,z}|X_3, S, Z, M, G^{(k+1)}\right) \propto \prod_{i\in G_i^{(k+1)}=g, Z_i=z} \frac{1_{(M_i=0)}+1_{(M_i=1)}e^{X_{3,i}\theta_{g,z}}}{1+e^{X_{3,i}\theta_{g,z}}},$$
$$\text{where } (G,Z) \in \{(LL,1),(LL,0),(LD,1),(DL,0)\}$$

C.4.3 Full Conditional Distributions of η_g

$$\eta_{LL}|Y, X_{1,LL}, S, Z, M, G^{(k+1)}, \sigma_{LL}^{2(k)} \sim N_p\left(\mu_{LL}^*, \sigma_{LL}^{2(k)}\left(X_{1,LL}^T X_{1,LL} + V_{LL}^{-1}\right)^{-1}\right)$$
$$\mu_{LL}^* = \left(X_{1,LL}^T X_{1,LL} + V_{LL}^{-1}\right)^{-1}\left(X_{1,LL}^T Y + V_{LL}^{-1}\mu_{LL}\right)$$

$$\eta_{LD}|Y, X_{1,LD}, S, Z, M, G^{(k+1)}, \sigma_{LD}^{2(k)} \sim N_p\left(\mu_{LD}^*, \sigma_{LD}^{2(k)}\left(X_{1,LD}^T X_{1,LD} + V_{LD}^{-1}\right)^{-1}\right)$$
$$\mu_{LD}^* = \left(X_{1,LD}^T X_{1,LD} + V_{LD}^{-1}\right)^{-1}\left(X_{1,LD}^T Y + V_{LD}^{-1}\mu_{LD}\right)$$

$$\eta_{DL}|Y, X_{1,DL}, S, Z, M, G^{(k+1)}, \sigma_{DL}^{2(k)} \sim N_p\left(\mu_{DL}^*, \sigma_{DL}^{2(k)}\left(X_{1,DL}^T X_{1,DL} + V_{DL}^{-1}\right)^{-1}\right)$$
$$\mu_{DL}^* = \left(X_{1,DL}^T X_{1,DL} + V_{DL}^{-1}\right)^{-1}\left(X_{1,DL}^T Y + V_{DL}^{-1}\mu_{DL}\right)$$

C.4.4 Full Conditional Distributions of σ_g^2

$$\eta_{LL}|Y, X_{1,LL}, S, Z, M, G^{(k+1)}, \eta_{LL}^{(k+1)} \sim IG\left(\nu_{LL} + \frac{n_{LL}}{2} + \frac{1}{2}, \omega_{LL}^*\right)$$
$$\omega_{LL}^* = \omega_{LL} + \frac{1}{2}\left[\left(\eta_{LL}^{(k+1)} - \mu_{LL}\right)^T V_{LL}^{-1}\left(\eta_{LL}^{(k+1)} - \mu_{LL}\right)\right.$$
$$\left. + \left(Y - X_{1,LL}\eta_{LL}^{(k+1)}\right)^T\left(Y - X_{1,LL}\eta_{LL}^{(k+1)}\right)\right]$$

$$\eta_{LD}|Y, X_{1,LD}, S, Z, M, G^{(k+1)}, \eta_{LD}^{(k+1)} \sim IG\left(\nu_{LD} + \frac{n_{LD}}{2} + \frac{1}{2}, \omega_{LD}^{*}\right)$$

$$\omega_{LD}^{*} = \omega_{LD} + \frac{1}{2}\left[\left(\eta_{LD}^{(k+1)} - \mu_{LD}\right)^{T} V_{LD}^{-1}\left(\eta_{LD}^{(k+1)} - \mu_{LD}\right) + \left(Y - X_{1,LD}\eta_{LD}^{(k+1)}\right)^{T}\left(Y - X_{1,LD}\eta_{LD}^{(k+1)}\right)\right]$$

$$\eta_{DL}|Y, X_{1,DL}, S, Z, M, G^{(k+1)}, \eta_{DL}^{(k+1)} \sim IG\left(\nu_{DL} + \frac{n_{DL}}{2} + \frac{1}{2}, \omega_{DL}^{*}\right)$$

$$\omega_{DL}^{*} = \omega_{DL} + \frac{1}{2}\left[\left(\eta_{LL}^{(k+1)} - \mu_{DL}\right)^{T} V_{DL}^{-1}\left(\eta_{DL}^{(k+1)} - \mu_{DL}\right) + \left(Y - X_{1,DL}\eta_{DL}^{(k+1)}\right)^{T}\left(Y - X_{1,DL}\eta_{DL}^{(k+1)}\right)\right]$$

Bibliography

Andersen, P. M., Borasio, G. D., Dengler, R., Hardiman, O., Kollewe, K., Leigh, P. N., Pradat, P.-F., Silani, V. & Tomik, B. (2007), 'Good practice in the management of amyotrophic lateral sclerosis: clinical guidelines. an evidence-based review with good practice points. ealsc working group', *Amyotrophic lateral sclerosis* **8**(4), 195–213.

Angrist, J. D., Imbens, G. W. & Rubin, D. B. (1996), 'Identification of causal effects using instrumental variables', *Journal of the American statistical Association* **91**(434), 444–455.

Anstrom, K. J. & Tsiatis, A. A. (2001), 'Utilizing propensity scores to estimate causal treatment effects with censored time-lagged data', *Biometrics* **57**(4), 1207–1218.

Austin, P. C. (2008), 'A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003', *Statistics in medicine* **27**(12), 2037–2049.

Austin, P. C. (2012), 'Generating survival times to simulate cox proportional hazards models with time-varying covariates', *Statistics in Medicine* **31**(29), 3946–3958.

Barnard, J., Frangakis, C. E., Hill, J. L. & Rubin, D. B. (2003), 'Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in new york city', *Journal of the American Statistical Association* **98**(462), 299–323.

Chio, A., Finocchiaro, E., Meineri, P., Bottacchi, E., Schiffer, D. et al. (1999), 'Safety and factors related to survival after percutaneous endoscopic gastrostomy in als', *Neurology* **53**(5), 1123–1123.

Chio, A., Galletti, R., Finocchiaro, C., Righi, D., Ruffino, M., Calvo, A., Di Vito, N., Ghiglione, P., Terreni, A. & Mutani, R. (2004), 'Percutaneous radiological gastrostomy: a safe and effective method of nutritional tube placement in advanced als', *Journal of Neurology, Neurosurgery & Psychiatry* **75**(4), 645–647.

Cole, S. R. & Hernán, M. A. (2008), 'Constructing inverse probability weights for marginal structural models', *American journal of epidemiology* **168**(6), 656–664.

Cox, D. R. et al. (1958), 'Planning of experiments', *Planning of Experiments.* .

D'Agostino, R. B. (1998), 'Tutorial in biostatistics: propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group', *Stat Med* **17**(19), 2265–2281.

Desport, J. C., Preux, P. M., Truong, C. T., Courat, L., Vallat, J. M. & Couratier, P. (2000), 'Nutritional assessment and survival in als patients', *Amyotrophic Lateral Sclerosis* **1**(2), 91–96.

Desport, J.-C., Torny, F., Lacoste, M., Preux, P.-M. & Couratier, P. (2006), 'Hypermetabolism in als: correlations with clinical and paraclinical parameters', *Neurodegenerative Diseases* **2**(3-4), 202–207.

Egleston, B. L., Scharfstein, D. O. & MacKenzie, E. (2009), 'On estimation of the survivor average causal effect in observational studies when important confounders are missing due to death', *Biometrics* **65**(2), 497–504.

Faraway, J. J. (2006), *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*, Vol. 66, Chapman & Hall.

Foster, E. M. (2003), 'Propensity score matching: an illustrative analysis of dose response', *Medical care* **41**(10), 1183–1192.

Frangakis, C. E. & Rubin, D. B. (1999), 'Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes', *Biometrika* **86**(2), 365–379.

Frangakis, C. E. & Rubin, D. B. (2002*a*), 'Principal stratification in causal inference.', *Biometrics* **58**(1), 21–29.

Frangakis, C. E. & Rubin, D. B. (2002*b*), 'Principal stratification in causal inference', *Biometrics* **58**(1), 21–29.

Frangakis, C. E., Rubin, D. B., An, M.-W. & MacKenzie, E. (2007), 'Principal stratification designs to estimate input data missing due to death', *Biometrics* **63**(3), 641–649.

Frumento, P., Mealli, F., Pacini, B. & Rubin, D. B. (2012), 'Evaluating the effect of training on wages in the presence of noncompliance, nonemployment, and missing outcome data', *Journal of the American Statistical Association* **107**(498), 450–466.

Gelinas, D. F. & Miller, R. G. (2000), 'A treatable disease: a guide to the management of amyotrophic lateral sclerosis', *Amyotrophic lateral sclerosis.* pp. 405–21.

Goyal, N. A. & Mozaffar, T. (2014), 'Respiratory and nutritional support in amyotrophic lateral sclerosis', *Current treatment options in neurology* **16**(2), 1–12.

Greenland, S. & Finkle, W. D. (1995), 'A critical look at methods for handling missing covariates in epidemiologic regression analyses', *American journal of epidemiology* **142**(12), 1255–1264.

Greenland, S., Pearl, J. & Robins, J. M. (1999), 'Causal diagrams for epidemiologic research', *Epidemiology* **10**(1), 37–48.

Gregory, S., Siderowf, A., Golaszewski, A. L. & McCluskey, L. (2002), 'Gastrostomy insertion in als patients with low vital capacity: respiratory support and survival', *Neurology* **58**(3), 485–487.

Guo, S. & Fraser, M. W. (2010), 'Propensity score analysis', *Statistical methods and applications* .

Hayden, D., Pauler, D. K. & Schoenfeld, D. (2005), 'An estimator for treatment comparisons among survivors in randomized trials', *Biometrics* **61**(1), 305–310.

Hernán, M. A. & Robins, J. M. (2006), 'Estimating causal effects from epidemiological data', *Journal of Epidemiology and Community Health* **60**(7), 578–586.

Hirano, K. & Imbens, G. W. (2004), 'The propensity score with continuous treatments', *Applied Bayesian modeling and causal inference from incomplete-data perspectives* pp. 73–84.

Holland, P. W. (1986), 'Statistics and causal inference', *Journal of the American Statistical Association* **81**(396), 945–960.

Hooker, G. (2010), Functional data analysis: A short course, *in* 'International Workshop on Statistical Modeling', Glasgow University.

Imai, K. & Van Dyk, D. A. (2004), 'Causal inference with general treatment regimes: Generalizing the propensity score', *Journal of the American Statistical Association* **99**(467), 854–866.

Imbens, G. W. (2000), 'The role of the propensity score in estimating dose-response functions', *Biometrika* **87**(3), 706–710.

Imbens, G. W. & Rubin, D. B. (1997), 'Bayesian inference for causal effects in randomized experiments with noncompliance', *The Annals of Statistics* pp. 305–327.

Jin, H. & Rubin, D. B. (2008), 'Principal stratification for causal inference with extended partial compliance', *Journal of the American Statistical Association* **103**(481), 101–111.

Jo, B. & Stuart, E. A. (2009), 'On the use of propensity scores in principal causal effect estimation', *Statistics in medicine* **28**(23), 2857–2875.

Joffe, M. M. & Rosenbaum, P. R. (1999), 'Invited commentary: Propensity scores', *American Journal of Epidemiology* **150**(4), 327–333.

Kasarskis, E. J., Berryman, S., Vanderleest, J. G., Schneider, A. R. & McClain, C. (1996), 'Nutritional status of patients with amyotrophic lateral sclerosis: relation to the proximity of death.', *The American journal of clinical nutrition* **63**(1), 130–137.

Katzberg, H. D. & Benatar, M. (2011), 'Enteral tube feeding for amyotrophic lateral sclerosis/motor neuron disease', *Cochrane Database Syst Rev* **1**.

Keisuke Hirano, Guido W. Imbens, G. R. (2003), 'Efficient estimation of average treatment effects using the estimated propensity scoreefficient estimation of average treatment effects fusing the estimated propensity score', *Econometrica* **71**(4), 1161–1189.

Kiernan, M. C., Vucic, S., Cheah, B. C., Turner, M. R., Eisen, A., Hardiman, O., Burrell, J. R. & Zoing, M. C. (2011), 'Amyotrophic lateral sclerosis', *The Lancet* **377**(9769), 942–955.

Körner, S., Hendricks, M., Kollewe, K., Zapf, A., Dengler, R., Silani, V. & Petri, S. (2013), 'Weight loss, dysphagia and supplement intake in patients with amyotrophic lateral sclerosis (als): impact on quality of life and therapeutic options', *BMC neurology* **13**(1), 84.

Leon, A. C., Mueller, T. I., Solomon, D. A. & Keller, M. B. (2001), 'A dynamic adaptation of the propensity score adjustment for effectiveness analyses of ordinal doses of treatment', *Statistics in medicine* **20**(9-10), 1487–1498.

Li, Y. P., Propert, K. J. & Rosenbaum, P. R. (2001), 'Balanced risk set matching', *Journal of the American Statistical Association* **96**(455), 870–882.

Little, R. J. (1995), 'Modeling the drop-out mechanism in repeated-measures studies', *Journal of the American Statistical Association* **90**(431), 1112–1121.

Little, R. J., Long, Q. & Lin, X. (2009), 'A comparison of methods for estimating the causal effect of a treatment in randomized clinical trials subject to noncompliance', *Biometrics* **65**(2), 640–649.

Little, R. J. & Rubin, D. B. (2000), 'Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches', *Annual review of public health* **21**(1), 121–145.

Little, R. J. & Yau, L. H. (1998), 'Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using rubin's causal model', *Psychological Methods* **3**, 147–159.

Long, Q., Little, R. J. & Lin, X. (2010), 'Estimating causal effects in trials involving multitreatment arms subject to non-compliance: a bayesian framework', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **59**(3), 513–531.

Lu, B. (2005), 'Propensity score matching with time-dependent covariates', *Biometrics* **61**(3), 721–728.

Mazzini, L., Corra, T., Zaccala, M., Mora, G., Del Piano, M. & Galante, M. (1995), 'Percutaneous endoscopic gastrostomy and enteral nutrition in amyotrophic lateral sclerosis', *Journal of neurology* **242**(10), 695–698.

McCulloch, C. E. (2006), *Generalized linear mixed models*, Wiley Online Library.

Miller, R. G., Rosenberg, J. A., Gelinas, D., Mitsumoto, H., Newman, D., Sufit, R., Borasio, G. D., Bradley, W. G., Bromberg, M. B., Brooks, B. R. et al. (1999), 'Practice parameter: The care of the patient with amyotrophic lateral sclerosis (an evidence-based review) report of the quality standards subcommittee of the american academy of neurology', *Neurology* **52**(7), 1311–1311.

Mitchell, J. D. & Borasio, G. D. (2007), 'Amyotrophic lateral sclerosis', *The lancet* **369**(9578), 2031–2041.

Mitsumoto, H., Davidson, M., Moore, D., Gad, N., Brandis, M., Ringel, S., Rosenfeld, J., Shefner, J., Strong, M., Sufit, R. et al. (2003), 'Percutaneous endoscopic gastrostomy (peg) in patients with als and bulbar dysfunction', *Amyotrophic Lateral Sclerosis* **4**(3), 177–185.

Muscaritoli, M., Kushta, I., Molfino, A., Inghilleri, M., Sabatelli, M. & Rossi Fanelli, F. (2012), 'Nutritional and metabolic support in patients with amyotrophic lateral sclerosis', *Nutrition* **28**(10), 959–966.

Ngo, S., Steyn, F. & McCombe, P. (2014), 'Body mass index and dietary intervention: Implications for prognosis of amyotrophic lateral sclerosis', *Journal of the Neurological Sciences* **340**(1), 5–12.

of Neurology ALS Work Group, A. A. (2012), 'Amyotrophic lateral sclerosis, performance measurement set'.

Orrell, R. W. (2007), 'Understanding the causes of amyotrophic lateral sclerosis', *New England Journal of Medicine* **357**(8), 822–823.

Paganoni, S., Deng, J., Jaffa, M., Cudkowicz, M. E. & Wills, A.-M. (2011), 'Body mass index, not dyslipidemia, is an independent predictor of survival in amyotrophic lateral sclerosis', *Muscle & nerve* **44**(1), 20–24.

Park, R., Allison, M., Lang, J. e., Spence, E., Morris, A., Danesh, B., Russell, R. & Mills, P. (1992), 'Randomised comparison of percutaneous endoscopic gastrostomy and nasogastric tube feeding in patients with persisting neurological dysphagia.', *British Medical Journal* **304**(6839), 1406.

Pastula, D. M., Coffman, C. J., Allen, K. D., Oddone, E. Z., Kasarskis, E. J., Lindquist, J. H., Morgenlander, J. C., Norman, B. B., Rozear, M. P., Sams, L. A. et al. (2009), 'Factors associated with survival in the national registry of veterans with als', *Amyotrophic Lateral Sclerosis* **10**(5-6), 332–338.

Procaccini, N. J. & Nemergut, E. C. (2008), 'Percutaneous endoscopic gastrostomy in the patient with amyotrophic lateral sclerosis: Risk vs benefit?', *Practical Gastroenterology* p. 24.

Radunović, A., Mitsumoto, H. & Leigh, P. N. (2007), 'Clinical care of patients with amyotrophic lateral sclerosis', *The Lancet Neurology* **6**(10), 913–925.

Ramsay, J., Hooker, G. & Graves, S. (2009), Introduction to functional data analysis, *in* 'Functional Data Analysis with R and MATLAB', Springer.

Ramsay, J., Wickham, H., Graves, S. & Hooker, G. (2011), 'fda: Functional data analysis', *R package version* **2**(6).

Robins, J. M. (1999), 'Association, causation, and marginal structural models', *Synthese* **121**(1), 151–179.

Robins, J. M., Hernán, M. Á. & Brumback, B. (2000), 'Marginal structural models and causal inference in epidemiology', *Epidemiology* **11**(5), 550–560.

Robins, J. M., Rotnitzky, A. & Zhao, L. P. (1994), 'Estimation of regression coefficients when some regressors are not always observed', *Journal of the American Statistical Association* **89**(427), 846–866.

Rosenbaum, P. R. & Rubin, D. B. (1983), 'The central role of the propensity score in observational studies for causal effects', *Biometrika* **70**(1), 41–55.

Rosenbaum, P. R. & Rubin, D. B. (1984), 'Reducing bias in observational studies using subclassification on the propensity score', *Journal of the American Statistical Association* **79**(387), 516–524.

Rosenbaum, P. R. & Rubin, D. B. (1985*a*), 'The bias due to incomplete matching', *Biometrics* pp. 103–116.

Rosenbaum, P. R. & Rubin, D. B. (1985*b*), 'Constructing a control group using multivariate matched sampling methods that incorporate the propensity score', *The American Statistician* **39**(1), 33–38.

Rotnitzky, A. & Robins, J. M. (1995), 'Semiparametric regression estimation in the presence of dependent censoring', *Biometrika* **82**(4), 805–820.

Rowland, L. P. & Shneider, N. A. (2001), 'Amyotrophic lateral sclerosis', *New England Journal of Medicine* **344**(22), 1688–1700.

Rubin, D. B. (1974), 'Estimating causal effects of treatments in randomized and nonrandomized studies', *Journal of Educational Psychology* **66**(5), 688–701.

Rubin, D. B. (1980), 'Randomization analysis of experimental data: The fisher randomization test (comment)', *Journal of the American Statistical Association* **75**(371), 591–593.

Rubin, D. B. (1990), 'Comment: Neyman (1923) and causal inference in experiments and observational studies', *Statistical Science* **5**(4), 472–480.

Rubin, D. B. (2004), 'Direct and indirect causal effects via potential outcomes*', *Scandinavian Journal of Statistics* **31**(2), 161–170.

Rubin, D. B. (2005), 'Causal inference using potential outcomes', *Journal of the American Statistical Association* **100**(469).

Rubin, D. B. & Thomas, N. (1996), 'Matching using estimated propensity scores: relating theory to practice', *Biometrics* pp. 249–264.

Rubin, D. B. et al. (2006), 'Causal inference through potential outcomes and principal stratification: application to studies with "censoring" due to death', *Statistical Science* **21**(3), 299–309.

Schwartz, S., Li, F. & Reiter, J. P. (2012), 'Sensitivity analysis for unmeasured confounding in principal stratification settings with binary variables', *Statistics in Medicine* **31**(10), 949–962.

Seeger, J. D., Williams, P. L. & Walker, A. M. (2005), 'An application of propensity score matching using claims data', *Pharmacoepidemiology and drug safety* **14**(7), 465–476.

Sekhon, J. (2008), 'Multivariate and propensity score matching software with automated balance optimization: the matching package for r', *Journal of Statistical Software, Forthcoming* .

Shaw, A. S., Ampong, M.-A., Rio, A., Al-Chalabi, A., Sellars, M. E., Ellis, C., Shaw, C. E., Leigh, N. P. & Sidhu, P. S. (2006), 'Survival of patients with als following institution of enteral feeding is related to pre-procedure oximetry: a retrospective review of 98 patients in a single centre', *Amyotrophic Lateral Sclerosis* **7**(1), 16–21.

Spataro, R., Ficano, L., Piccoli, F. & La Bella, V. (2011), 'Percutaneous endoscopic gastrostomy in amyotrophic lateral sclerosis: effect on survival', *Journal of the neurological sciences* **304**(1), 44–48.

Tanner, M. A. & Wong, W. H. (1987), 'The calculation of posterior distributions by data augmentation', *Journal of the American statistical Association* **82**(398), 528–540.

Traxinger, K., Kelly, C., Johnson, B. A., Lyles, R. H. & Glass, J. D. (2013), 'Prognosis and epidemiology of amyotrophic lateral sclerosis analysis of a clinic population, 1997–2011', *Neurology: Clinical Practice* **3**(4), 313–320.

VanderWeele, T. J. (2008), 'Simple relations between principal stratification and direct and indirect effects', *Statistics & Probability Letters* **78**(17), 2957–2962.

Verschueren, A., Monnier, A., Attarian, S., Lardillier, D. & Pouget, J. (2009), 'Enteral and parenteral nutrition in the later stages of als: an observational study', *Amyotrophic Lateral Sclerosis* **10**(1), 42–46.

Yau, L. H. Y. & Little, R. J. (2001), 'Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed', *Journal of the American Statistical Association* **96**(456).

Zhang, J. L. & Rubin, D. B. (2003), 'Estimation of causal effects via principal stratification when some outcomes are truncated by "death"', *Journal of Educational and Behavioral Statistics* **28**(4), 353–368.

Zhang, J. L., Rubin, D. B. & Mealli, F. (2008), 'Evaluating the effects of job training programs on wages through principal stratification', *Advances in Econometrics* **21**, 117–145.

Zhang, J. L., Rubin, D. B. & Mealli, F. (2009), 'Likelihood-based analysis of causal effects of job-training programs using principal stratification', *Journal of the American Statistical Association* **104**(485).

Zhao, H. & Tsiatis, A. A. (1997), 'A consistent estimator for the distribution of quality adjusted survival time', *Biometrika* **84**(2), 339–348.

Zhao, S., van Dyk, D. A. & Imai, K. (2012), 'Causal inference in observational studies with non-binary treatments'.

Zigler, C. M., Watts, K., Yeh, R. W., Wang, Y., Coull, B. A. & Dominici, F. (2013), 'Model feedback in bayesian propensity score estimation', *Biometrics* .