Abstract

In observational clinic registries, time to treatment is often of interest, but treatment can be given at any time during follow-up and there is no structure or intervention to ensure regular clinic visits for data collection. To address these challenges, we introduce the time-dependent propensity process as a generalization of the propensity score. We show that the propensity process balances the entire time-varying covariate history which cannot be achieved by existing propensity score methods and that treatment assignment is strongly ignorable conditional on the propensity process. We develop methods for estimating the propensity process using observed data and for matching based on the propensity process. We illustrate the propensity process method using the Emory Amyotrophic Lateral Sclerosis (ALS) Registry data.

Keywords: Balancing Score; Generalized Propensity Score; Propensity Process; Propensity Score; Observational Registry; Time-Varying Covariates
1 Introduction

Amyotrophic lateral sclerosis (ALS) is a rare progressive neurological 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, with a very poor prognosis. Currently, there is no cure for ALS and clinical care is generally limited to treating secondary infections and palliative care, such as surgically inserting a percutaneous endogastrostomy (PEG) tube to provide enteral nutrition for individuals having difficulty swallowing (Procaccini and Nemergut, 2008). Our objective is to assess the effect of inserting a PEG feeding tube on preventing weight loss. PEG insertion is an individual decision and one that must be made while the individual is strong enough to proceed with surgery. Hence, a randomized controlled trial to study the effect of PEG would be implausible. We develop new methods to evaluate PEG using data from the Emory ALS Clinic registry.

Let \( T \) denote the continuously-defined time of PEG insertion for a randomly selected patient from the population. The observed outcome \( Y \) is collected at or just after a fixed point at time \( L \), which consequently restricts the time of PEG insertion. If subjects were randomly assigned to receive PEG prior to \( L \) and randomly assigned to treatment times, then both treatment effect and dose-response curve could be estimated using standard methods. However, treatment assignment depends on patient characteristics and confounds the effect of treatment on outcome. To remove confounding associated with covariate imbalance among treatment levels, we rely on the general concept of the propensity score (Rosenbaum and Rubin, 1983; Rubin and Thomas, 1996).

When treatment assignment is binary, the propensity score (Rosenbaum and Rubin, 1983) is defined as probability of receiving a treatment given a set of observed variables. Generalizations of the propensity score as a balancing score have been investigated in various settings (Hirano and Imbens, 2004; Imai and Van Dyk, 2004; Hansen, 2008; Allen and Satten, 2011; Hu et al., 2014). For continuously-defined treatment levels, Hirano and Imbens (2004) proposed a direct translation of the propensity score by replacing the conditional probability mass function with the conditional density function of treatment assignment given covariates, known as a generalized propensity score (GPS); while
this approach leads to as many propensity scores as there are levels of the treatment it uses only one single score at a time. Although Imai and Van Dyk (2004) similarly found that the conditional density function of treatment assignment given covariates could serve as a propensity score, they noted potential limitations of this approach and suggested instead using the linear predictor in regression models or other summary statistic that are of finite dimension. When treatment assignment occurs over time as in the case where an individual chooses to receive PEG insertion or not, we must allow for the possibility of time-dependent confounding. To this end, let $X_t$ denote a set of $p$-dimensional time-dependent covariates at time $t$ and $\mathcal{X}_t = \{X_s, 0 \leq s \leq t\}$ denote the history of covariates up to time $t$. Then, the probability of treatment assignment at time $t$ given the covariate history up to time $t$ is

$$f(t \mid \mathcal{X}_t) = \lim_{\epsilon \to 0} \epsilon^{-1} P(t \leq T < t + \epsilon \mid \mathcal{X}_t),$$

(1)

where $f(t \mid \mathcal{X}_t) = h(t \mid X_t) \exp \left\{- \int_0^t h(s \mid X_s) \, ds \right\}$ and the hazard function is

$$h(t \mid X_t) = \lim_{\epsilon \to 0} \epsilon^{-1} P(t \leq T < t + \epsilon \mid T \geq t, X_t).$$

(2)

Because $h(t \mid X_t)$ uniquely parameterises $f(t \mid \mathcal{X}_t)$, either model (1) or model (2) may be regarded as a legitimate treatment assignment model for continuous treatment with time-independent or time-dependent confounding (Li et al., 2001; Lu, 2005).

Of note, $f(t \mid \mathcal{X}_t)$ is a function of the entire covariate history $\mathcal{X}_t$, whereas the hazard function $h(t \mid X_t)$ is a function of $X_t$ only. This subtle, yet important difference can lead to difficulties when extending methods proposed by Imai and Van Dyk (2004) and Hirano and Imbens (2004) to time-dependent confounding via standard hazard modeling. In addition, both Li et al. (2001) and Lu (2005) used the hazard function $h(t \mid X_t)$ as a GPS for matching which allows for balancing $X_t$ at the time of treatment in a matched set. However, they did not establish the strong ignorability of treatment assignment given their time-dependent GPS; this property does not hold if $Y$ is associated with $\mathcal{X}_t$ rather than just $X_t$, in which case their proposed procedures may not lead to valid causal inference. Additionally, their proposed methods are only applicable to studies with data routinely collected at regular intervals, which is often not true in clinical registries.
We propose the propensity process to correct for confounding in observational studies by balancing the covariate history $X_t$. After the propensity process is estimated, bias-corrected data analyses can be achieved through matching or stratification (Rosenbaum and Rubin, 1983). Establishing formally the theoretical properties of the propensity process for time-independent confounding requires different arguments than those presented in Imai and Van Dyk (2004).

2 Methods

2.1 Notation and Assumptions

Our framework is constructed through potential outcomes (Rubin, 2005). For $t \in [0, L)$, we define $U_t = T \wedge t$ as the treatment time restricted to time $t$ and $U = T \wedge L$ as the treatment time restricted to time $L$, where $a \wedge b$ denotes the minimum of $a$ and $b$. Let $T_t = \{(0, t), t+\}$ define the set of potential treatment times restricted to $t$, $t \in [0, L)$, where $t+$ means that a patient did not receive PEG treatment before $t$. Let $Y_t^*$ be the potential outcome if a subject received PEG treatment at time $t$, $t \in [0, L)$, and $Y_{t+}^*$ be the potential outcome if a subject did not receive PEG treatment in the interval $[0, t)$. It follows that $Y_{L+}^*$ denotes the potential outcome if a subject did not receive PEG treatment in the interval $[0, L)$. We also define the treatment-free potential covariate process $X_t^*$, $t \leq L$. Then, the set of potential outcomes and treatment-free potential covariate process for a randomly selected subject from the population is $\{Y_s^*, X_s^*, s \in T_t\}$ when treatment time is restricted at $t$, $t \in [0, L)$. In contrast, the observed data are $(Y, U, X_U)$, where the observed outcome $Y = Y_U^*$, and the observed covariate history $X_U = X_U^*$.

Given $\theta_t = h(t \mid X_t^*)$, we define the propensity process as the sample path of the hazard function from baseline to time $t$, i.e.,

$$\Theta_t = \{\theta_s = h(s \mid X_s^*), 0 \leq s \leq t\}, \quad (3)$$

noting that $\Theta_t$ is dependent on $X_t^*$. As $X_t^*$ is observable only up to $U$, $\Theta_t$ is estimable only up to $U$. While this concept seems similar to the propensity function (Imai and Van Dyk, 2004).
The distinguishing factor of the propensity process is that \( \Theta_t \) depends on \( t \) and is of infinite dimension and \( \Theta_L \) cannot be fully estimated for subjects receiving PEG before \( L \), whereas the propensity function in Imai and Van Dyk (2004) only allows for incorporation of time-independent covariates and can be estimated for all subjects.

In our framework, we make two assumptions.

**Assumption 1 (Stable unit treatment value assumption)** The distributions of potential outcomes for different subjects are independent of one another.

**Assumption 2 (Strong Ignorability)** For every \( t \in [0, L) \), \( \text{pr}(U_t \in A \mid Y^*_s, X^*_t) = \text{pr}(U_t \in A \mid X^*_t) \) and \( \text{pr}(U_t \in A \mid X^*_t) > 0 \) for all \( s \in T_t, X^*_t, \) and \( A \subseteq T_t \).

Assumption 1 is a common assumption in causal inference. However, our Assumption 2 is defined for each time point \( t \) and differs from the standard strong ignorability of treatment assignment assumption used in earlier work for balancing scores. One implication of Assumption 2 is that, conditional on the treatment-free history \( X^*_t \), receiving treatment at \( t \) or not is independent of the set of potential outcomes, allowing us to model treatment assignment without conditioning on potential outcomes.

### 2.2 Main results

We establish the large-sample results of the propensity process assuming that the true propensity process is known along the lines of Rosenbaum and Rubin (1983) and Imai and Van Dyk (2004).

**Proposition 1** \( U \) is conditionally independent of treatment-free covariate history \( X^*_t \) given \( \Theta_L \), where \( X^*_t \) and \( \Theta_L \) are the entire treatment-free covariate history and propensity process, respectively.

Proposition 1 establishes \( \Theta_L \) as a balancing functional that balances the entire covariate history. Proposition 1 requires that \( \Theta_L \) is known or can be estimated in the entire domain \([0, L)\). In practice, however, we can only observe the covariate process \( X^*_U \) and hence estimate \( \Theta_U \). Proposition 2 establishes the balancing property for every given time point \( t \) in \([0, L)\).
**Proposition 2** For every \( t \in [0, L) \), \( U_t \) is conditionally independent of treatment-free covariate history \( X_t^* \) given \( \Theta_t \), where \( X_t^* \) and \( \Theta_t \) are the treatment-free covariate history and propensity process through time \( t \), respectively.

When \( t = U \) in Proposition 2, we have that \( U \) is independent of treatment-free covariate history \( X_U^* \) given \( \Theta_U \), where \( X_U^* = X_U \) is observable and hence \( \Theta_U \) is estimable.

**Theorem 1** For every \( t \in [0, L) \), \( pr(U_t \in A \mid Y_s^*, \Theta_t) = pr(U_t \in A \mid \Theta_t) \) for all \( s \in T_t, \Theta_t, \) and \( A \subseteq T_t \).

When \( t = U \) in Theorem 1, we have that \( U \) is independent of potential outcomes given \( \Theta_U \), where \( \Theta_U \) is estimable. Several remarks are in order. First, in § 3.1 we suggest modeling the hazard function in (2) through the proportional hazards model (4); one could also use other model formulations for (2) and the results in Propositions 1–2 and Theorem 1 would still apply. Second, Proposition 2 and Theorem 1 provide justifications for matching a subject treated at \( t \) with an eligible control subject untreated at \( t \) based on the propensity process up to \( t \). It follows that each matched pair would have the same distribution for the covariate process up to \( t \) and their potential outcomes are independent of their treatment assignments, allowing for valid causal inference. Third, our Proposition 2 is similar in spirit to Proposition 1 in Lu (2005) but is more general in the sense that the propensity process balances the entire covariate history up to \( t \) not just the covariates measured at \( t \). In addition, Lu (2005) did not establish the strong ignorability of treatment assignment given propensity scores similar to our Theorem 1. Proofs for Propositions 1–2 and Theorem 1 are given in the Appendix.

3 Implementation and Practical Considerations

3.1 Interpolated Propensity Processes

In practice, the propensity process \( \Theta_U \) must be estimated from the observed data. The challenge for estimating the propensity process is that we may not observe the complete treatment-free covariate process \( X_U^* \) on \([0, U]\); rather, we only get to observe the covariate
process at a coarse set of discrete time points as is the case in the motivating ALS study. Here, we propose to borrow strength across subjects in the study sample by modeling each time-dependent covariate as a random curve over time via nonlinear mixed effects models. This allows a predictive curve to be estimated for the entire treatment-free covariate process for each subject.

First, suppose we parameterize the hazard function in (2) through Cox’s proportional hazards model and define the propensity process through the linear predictor,

\[ h(t \mid X_t; \beta) = h_0(t) \exp(\beta^T X_t), \quad \Theta_t = \{ \theta_s = \beta^T X_s^*, 0 \leq s \leq t \}, \]

where \( h_0(t) \) is the unspecified baseline hazard function. Next, write the observed treatment-free covariate history for the \( i \)-th subject and \( k \)-th covariate as \( X_{ik} = (X_{i1k}, \ldots, X_{im_k}) \), with time-dependent covariate \( X_{ijk} \) measured at time \( t_{ij} \). We note that the observation times \( (t_{ij}, j = 1, \ldots, m_i) \) may be different for each subject but are assumed to be the same for all covariates within a subject. Then, for each time-dependent covariate, we fit the model,

\[ X_{ijk} = b_k^T(t_{ij})\gamma_k + b_k^T(t_{ij})\alpha_{ik} + \epsilon_{ijk}, (i = 1, \ldots, n; j = 1, \ldots, m_i; k = 1, \ldots, p), \]

where \( \epsilon_{ijk} \) are independent, mean-zero random errors. To provide greater flexibility in modeling the covariate process over time, we use spline-type models (Ruppert et al., 2003) in (5) where \( b(\cdot) \) denotes a set of basis functions and \( \gamma_k \) and \( \alpha_{ik} \) are regression coefficients corresponding to the basis functions for the fixed and random effects, respectively. The interpolated treatment-free \( \hat{X}_t \) can be obtained from model (5) by replacing regression coefficients \( \gamma_k \) and \( \alpha_{ik} \) with their estimates \( \hat{\gamma}_k \) and \( \hat{\alpha}_{ik} \), respectively. Then the estimated propensity process \( \hat{\Theta}_U \) can be obtained from (4) by plugging in the interpolated \( \hat{X}_U \) and \( \hat{\beta} \), where \( \hat{\beta} \) is the estimated regression coefficient vector in the Cox proportional hazards model.

### 3.2 Matching

The use of matched analyses based on propensity scores for testing causal null hypotheses has been advocated by several other authors; for example, see Rosenbaum and Rubin.
Matching can be performed by minimizing the integrated squared error between the estimated propensity process $\hat{\Theta}_t$ of a subject who received PEG treatment at time $t$ and that of each eligible control with $U > t$. To accomplish this task, we implement a sequential matching algorithm. We start by ordering chronologically subjects according to their time of PEG treatment or censoring, namely $U$. Set the matched pair counter to $m = 1$ and select the subject with the smallest time to PEG treatment, say subject $i_1$. Define the integrated squared difference in interpolated propensity processes between $i_1$ and $l$ as $Q(i_1, l) = I(T_{i_1} \leq L) \int_0^{T_{i_1}} (\hat{\theta}_{i_1,t} - \hat{\theta}_{l,t})^2 dt$, for all subjects $l$ in the set of $n - 1$ eligible controls $C_1 = \{l \mid l = 1, \ldots, n, \ l \neq i\}$. The matched control for $i_1$ is the nearest neighbor in interpolated propensity processes among eligible controls, i.e., $\text{argmin}_{l \in C_1} Q(i_1, l)$. Increment the matched pair counter by one to $m = 2$ and select the subject with the smallest time to PEG treatment, say $i_2$, excluding the two subjects in the first matched pair. Therefore, the set of eligible controls, say $C_2$, contains $n - 3$ subjects: all $n$ subjects less the two subjects in the first matched pair and $i_2$. The matched control for $i_2$ is the nearest neighbor in interpolated propensity processes among the set of eligible controls, $\text{argmin}_{l \in C_2} Q(i_2, l)$. Increment the matched pair counter by one and continue until all treated individuals are matched or until there are no suitable controls available for matching.

4 Analysis of the ALS Registry Data

Using a data set from the Emory ALS Registry, we assess the association of PEG treatment with the change in body mass index (BMI) from baseline to 18 months, i.e., $L = 18$ months. The data set includes 240 patients who survived past $L$ and had at least one clinic visit between baseline and $L$. The patients who received PEG did so after their first clinic visit. The timing of recommending PEG by the physician involved many factors and the final decision to have PEG was made by each patient. We model treatment assignment through the proportional hazards model including the following covariates. The baseline risk factors are age at diagnosis, sex, site of onset of disease, negative inspiratory force, and time from diagnosis to the first clinic visit. Two time-varying covariates are forced vital
capacity and body mass index, which may not be measured at every clinic visit for every patient. Each time-varying covariate is modeled over time using the mixed model (5), where polynomial spline basis functions are used. The estimated curves are used to interpolate the covariate values needed for estimating the propensity process based on (4).

We compare three alternative approaches to the proposed propensity process. First, a naïve analysis compares all treated individuals to those who are untreated prior to \( L \). The second approach is the propensity function (Imai and Van Dyk, 2004) that uses baseline risk factors \( X_0 \) only in the treatment assignment model (4), where \( \theta_0 = \beta^T X_0 \) defines the propensity function. The third approach is the interpolated generalized propensity score, which uses the interpolated treatment-free \( \hat{X}_t \) defined in §3.1 to obtain the GPS for each subject in the spirit of Lu (2005), noting that \( X_t \) may not be observed at time \( U \) for a subject and its eligible controls as defined in §3.2. The same sequential matching algorithm in §3.2 is used for all propensity score methods. Our matching algorithm resulted in \( M = 74 \) pairs for the analysis using the propensity function and \( M = 76 \) pairs for both analyses using the generalized propensity score and propensity process.

Following Li et al. (2001) and Lu (2005), we assess balance of covariates by examining Type I errors from a log-rank test of the effect of the covariate on time to treatment, one covariate at a time. In the matched analyses, this model is stratified by the \( M \) matched pairs. As shown in Table 1, prior to matching, balance is not achieved. While other methods improve covariate balance, they do not balance all covariates. However, matching using the propensity process results in balance across all covariates. This indicates that the propensity process outperforms the baseline propensity function or interpolated GPS in terms of balancing covariates and there may be residual confounding after matching by the other propensity score methods.

After matching, we test the causal null hypothesis that the mean potential outcome is the same whether a patient received PEG treatment at time \( t \) versus PEG treatment at some time after \( t \) or untreated by \( L \), which can be written as \( H_0: E(Y_{t+}^*) = E(Y_{s+}^*) \) for all \( t < s \leq L \). We test this hypothesis by a Wilcoxon signed rank test on matched pairs for all the matched analyses. The Wilcoxon rank sum test is used for hypothesis testing in the
Table 1: Covariate balance before and after matching

| Covariate                  | Prior to Matching | Propensity Function | Generalized Propensity Score | Propensity Process |
|----------------------------|-------------------|---------------------|------------------------------|--------------------|
| Body mass index            | 0.277             | 0.245               | 0.986                        | 0.991              |
| Forced vital capacity      | 0.764             | 0.539               | 0.201                        | 0.317              |
| Negative inspiratory force | 0.151             | 0.022               | 0.016                        | 0.704              |
| Age                        | 0.162             | 0.718               | 0.378                        | 0.195              |
| Sex                        | 0.577             | 0.695               | 0.002                        | 0.706              |
| Site                       | 0.001             | 0.003               | 1.000                        | 0.341              |
| Time from diagnosis        | 0.676             | 0.633               | 0.033                        | 0.854              |

naïve analysis. Table 2 presents the median difference in BMI change at 18 months and p-value of the Wilcoxon test for each approach. The propensity process matched analysis suggests a protective effect of PEG on BMI, whereas the other three methods all show effects that are attenuated towards 0 and are not statistically significant.

Table 2: Results in the data analysis

| Covariate                  | Median Difference | P-value |
|----------------------------|-------------------|---------|
| Naïve                      | 0.035             | 0.673   |
| Propensity Function        | 0.030             | 0.466   |
| Generalized Propensity Score | 0.360             | 0.453   |
| Propensity Process         | 0.830             | 0.022   |

5 Discussion

Compared to the existing propensity score methods, the propensity process offers the advantage of balancing time-varying covariate history from baseline to time of treatment. A key component of this approach is the interpolation of covariate curves. We propose to use nonlinear mixed models to provide flexibility for modeling covariate history, though there
must be enough individual longitudinal data collected to estimate these curves, which is a potential limitation in settings with sparsely collected longitudinal data. However, data interpolation may not be needed in settings such as critical care in intensive care units where time series data including heart rate and blood pressure are continuously recorded (Lehman et al., 2013).

In our data analysis, we use a straightforward approach for hypothesis testing after matching. Future extensions may include conditional likelihood methods for estimating treatment effects based on matched pairs/sets and methods for stratification and covariate adjustment using the propensity process. Additionally, our analysis excludes individuals who died prior to \( L \) in order to avoid complications due to censoring by death (Rubin et al., 2006; Zhang and Rubin, 2003), which could be addressed in future extensions such that no such exclusion is necessary.

Acknowledgements

The authors thank Dr. Jonathan Glass and Ms. Meraida Polak at the Emory ALS Center for providing the ALS data and Dr. Xin Qi at the Georgia State University for helpful comments.

Appendix: Proofs of Propositions 1, 2, and Theorem 1

Proof of Propositions 1 and 2: We prove Propositions 1 and 2 based on the treatment assignment model defined in (1) and (2). Given \( \theta_t = h(t \mid X_t^*) \),

\[
\begin{align*}
f(t \mid X_t^*, \Theta_t) & = f(t \mid X_t^*) \\
& = h(t \mid X_t^*) \exp \left\{ - \int_0^t h(s \mid X_s^*) \, ds \right\} \\
& = \theta_t \exp \left\{ - \int_0^t \theta_s \, ds \right\} \\
& = f(t \mid \Theta_t), \text{ for all } t \in [0, L),
\end{align*}
\]
where the first equality is due to the fact that \( \Theta_t \) is redundant given \( X^*_t \). It follows from integrating both sides in \([0, L]\) that \( \Pr(T \geq L \mid X^*_L, \Theta_L) = \Pr(T \geq L \mid \Theta_L) \). The result in Proposition 1 follows immediately, i.e., \( U_t \) is conditionally independent of \( X^*_L \) given \( \Theta_L \). Along similar lines, we can prove the result in Proposition 2, i.e., \( U_t \) is conditionally independent of \( X^*_L \) given \( \Theta_t \) for all \( t \in [0, L] \).

**Proof of Theorem 1:** For every \( t \in [0, L] \), all \( s \in T_t \), \( \Theta_t \), and \( \mathcal{A} \subseteq T_t \),

\[
\Pr(U_t \in \mathcal{A} \mid Y^*_s, \Theta_t) = E\{\Pr(U_t \in \mathcal{A} \mid Y^*_s, X^*_t) \mid Y^*_s, \Theta_t\} \tag{7}
\]

\[
= E\{\Pr(U_t \in \mathcal{A} \mid X^*_t) \mid Y^*_s, \Theta_t\} \tag{8}
\]

\[
= E\{\Pr(U_t \in \mathcal{A} \mid X^*_t, \Theta_t) \mid Y^*_s, \Theta_t\} \tag{9}
\]

\[
= E\{\Pr(U_t \in \mathcal{A} \mid \Theta_t) \mid Y^*_s, \Theta_t\} \tag{10}
\]

\[
= \Pr(U_t \in \mathcal{A} \mid \Theta_t) \tag{11}
\]

Let \( \sigma(Y^*_s, \Theta_t) \) and \( \sigma(Y^*_s, X^*_t) \) denote the \( \sigma \)-field generated by \((Y^*_s, \Theta_t)\) and \((Y^*_s, X^*_t)\), respectively. By the definition of \( \Theta_t \) in (3), we have \( \sigma(Y^*_s, \Theta_t) \subseteq \sigma(Y^*_s, X^*_t) \) and then (7) follows immediately (cf. Billingsley, 2008, Theorem 34.4). (8) is due to Assumption 2 while (9) follows from the fact that \( \Theta_t \) is redundant given \( X^*_t \). The fourth expression (10) is due to Proposition 2 while (11) follows from (6), i.e., \( \Pr(U_t \in \mathcal{A} \mid \Theta_t) \) is independent of \( Y^*_s \).

**References**

Allen, A. S. and Satten, G. A. (2011). Control for confounding in case-control studies using the stratification score, a retrospective balancing score. *American journal of epidemiology*, 173(7):752–760.

Billingsley, P. (2008). *Probability and measure*. John Wiley & Sons.

Hansen, B. B. (2008). The prognostic analogue of the propensity score. *Biometrika*, 95(2):481–488.

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

Hu, Z., Follmann, D. A., and Wang, N. (2014). Estimation of mean response via the effective balancing score. Biometrika, page asu022.

Imai, K. and 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.

Lehman, L.-w. H., Nemati, S., Adams, R. P., Moody, G., Malhotra, A., and Mark, R. G. (2013). Tracking progression of patient state of health in critical care using inferred shared dynamics in physiological time series. In Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE, pages 7072–7075. IEEE.

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

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

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

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

Rubin, D. B. (2005). Causal inference using potential outcomes. Journal of the American Statistical Association, 100(469):322–331.

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

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric regression*. Number 12. Cambridge university press.

Zhang, J. L. and 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.