Estimation of Dose-Response Functions for Longitudinal Data

Erica E M Moodie*  David A. Stephens†

*McGill University, erica.moodie@mcgill.ca
†McGill University, d.stephens@math.mcgill.ca

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

http://biostats.bepress.com/cobra/art32

Copyright ©2007 by the authors.
Estimation of Dose-Response Functions for Longitudinal Data

Erica E M Moodie and David A. Stephens

Abstract

In a longitudinal study of dose-response, the presence of confounding or non-compliance compromises the estimation of the true effect of a treatment. Standard regression methods cannot remove the bias introduced by patient-selected treatment level, that is, they do not permit the estimation of the causal effect of dose. Using an approach based on the Generalized Propensity Score (GPS), a generalization of the classical, binary treatment propensity score, it is possible to construct a balancing score that provides a more meaningful estimation procedure for the true (unconfounded) effect of dose. Previously, the GPS has been applied only in a single interval setting. In this paper, we extend the GPS methodology to the longitudinal setting. The methodology is applied to simulated data and two real data sets; first, we study the Riesby depression data, and secondly we present analysis of a recent study, the Monitored Occlusion Treatment of Amblyopia Study (MOTAS), which investigated the dose-response relationship in an ophthalmological setting between occlusion and improvement in visual acuity. The MOTAS study was revolutionary as it was the first to accurately measure occlusion dose received by the child.
1. Introduction

In observational studies of the efficacy of a treatment, there is the potential for bias in the estimation of the treatment effect whenever the treatment dose level is influenced by subject-specific covariates. Also, randomized trials, particularly those where treatment is given over time in several treatment intervals, must contend with partial or total non-compliance, which effectively renders the trial observational. Statistical analyses in the face of non-compliance have often relied on intention-to-treat or as-treated analyses, which respectively ignore the dose actually received or do not account for the informative nature of non-compliance. The aim of this paper is to provide a framework for examining the effect of treatment given over time with incomplete adherence or at a patient-controlled level.

1.1. Longitudinal Observational Dose-Response Studies

Two approaches may be of interest when analyzing longitudinal data. The first assumes a repeated measures structure, with each participant providing repeated time-varying covariate data over a number of intermediate measurements as well as a final, end-of-study health measurement, and where the response is taken to be the vector of health outcome measurements. The second focuses on the by-interval changes, taking as the response the vector of changes in health outcomes between successive measurements. Where a by-interval approach provides a meaningful response, its use may simplify the analysis by reducing or even removing any serial correlation which may be present in the data. Using the vector of differences in response also has the advantage of allowing the analyst to fit a common curve in all intervals (for example, a linear dose-response with a common intercept as well as slope). Arjas and Parner (2004) provide a recent review of the main statistical issues that arise when estimating causal effects from observational longitudinal data.

In this paper, we focus on a by-interval approach, and estimate the cross-sectional effect of dosing by examining the change in response that may be attributed to a unit of dose in an interval. With data in this form, a regression approach may at first seem to provide a reasonable basis for analysis. This strategy, however, takes no account of subject-controlled treatment level, which we interpret as non- or partial compliance, which is a potential source of bias in the estimation of treatment effect. In this paper, we develop methods that address issues of confounding and non-compliance using a balancing score approach based on the Generalized Propensity Score (GPS) (Imbens, 2000; Hirano and Imbens, 2004) for a continuous treatment that controls for sources of such bias. The GPS has received relatively little attention in statistical circles; however, see Flores (2004) for an economics application.

As noted by Rubin (2007), observational studies and randomized trials are part of a con-
tinueum, due to potential confounding of the dose received in randomized studies due to non-compliance. In the context of randomized trials, causal methods have been proposed which are based on principal stratification by the compliance score (Follman, 2000; Frangakis and Rubin, 1999, 2002); see Joffe et al. (2003) for a comprehensive explanation and discussion. In such an approach, a compliance score, a predictive model for compliance with assigned treatment given baseline covariates, is used (frequently as a matching variable) to estimate the complier average causal effect (CACE), that is, the effect treatment among individuals who would comply with a prescribed treatment dose. Adjustments for compliance based only on pre-treatment variables will typically be unable to identify the effect of a treatment taken over time, as compliance may vary as a function of response to treatment. Furthermore, as we will demonstrate, a univariate scoring approach is not capable of accounting for all possible forms of confounding in the longitudinal setting.

Causal methods for estimating treatment effects from repeated measures data are available but not readily implemented for continuous doses. Neugebauer and van der Laan (2006), for example, propose a G-computation procedure. However the procedure is computationally expensive and so the authors discretized the continuous treatment. Another causal procedure for repeated measures data uses Marginal Structural Models (MSMs) (Robins et al., 2000; Hernán et al., 2000). MSMs use inverse weighting by the treatment mechanism model to remove bias and estimate the marginal effect of time-varying treatment regimes on a univariate response. MSMs were originally designed to estimate the effects of static (not covariate-adapted) time-varying treatments, but were recently proposed as a method to compare dynamic regimes which may have been found by other estimation procedures (Hernán et al., 2006) and further extended to estimate the optimal dynamic regime Bembom and van der Laan (2007). Marginal Structural Models can be used for continuous treatments, although this is not common in practice, in much the same way that the GPS is an under-recognized and under-used tool the single time-point setting. In a single interval, the MSM approach is very closely related to the Generalized Propensity Score analysis procedure.

Finally, research into estimating dynamic treatment regimes – that is, finding the best covariate-specific treatments – has been a particularly active area in causal inference for longitudinal data. Murphy (2003) and Robins (2004) published seminal material on methods for estimating optimal treatment rules, which are reviewed by Moodie et al. (2007).

1.2. Objectives and Structure of Paper

In this paper, we extend a causal modelling approach to account for the within-person correlation of responses, the within-person correlation of doses, and the potential confounding of the cross-sectional effect of dose on response by previous doses.
The paper is structured as follows: Section 2. introduces notation and the Generalized Propensity Score methodology. Section 3. extends the balancing score approach to incorporate the complexities of a repeated measures structure with time-varying covariates. The repeated measures GPS is compared with traditional regression methods and the univariate GPS via simulation in section 4.. Section 4.3. provides an example of GPS analyses applied to a familiar data set, the Reisby et al. (1977) study of depressed patients treated using the drug imipramine. Section 5. discusses and concludes.

2. A Balancing Score Approach to Estimating a Dose-Response Relationship

To ascertain the true effect of dose, a causal analysis which accounts for the potential confounding between dose levels and other measured covariates is desirable. One tool used to account for possible confounding relationships between occlusion treatment and other covariates is the Generalized Propensity Score (Imbens, 2000; Imai and Van Dyk, 2004; Hirano and Imbens, 2004), a variable that can be used for stratification or explicitly in a regression analysis that removes bias in the estimation of the treatment effect. A regression which includes the GPS does not provide a parameter that may be interpreted causally (see the discussion below), however it can be used to obtain estimates of the potential response to a dose which does have a causal interpretation. The balancing property of the GPS may provide greater confidence that the analysis has successfully eliminated potential bias due to (measured) confounding.

2.1. The Generalized Propensity Score

In the following formulation of the Generalized Propensity Score model, we presume that the dose/response pairs are conditionally independent, that is, we ignore the repeated measures nature of the data. In Section 3., we extend the framework to one suitable for the analysis of longitudinal observational data.

Suppose that we have collected data repeatedly on $N$ individuals, so that $n_i$, $i = 1, ..., N$ measurements are available. We denote the total number of data points $n = \sum_{i=1}^{N} n_i$, the response $Y$, the treatment dose $D$, and other potentially confounding covariates $X$; we denote observed values of these random variables $y, d$ and $x$ respectively. We define $\mathcal{D}$ to be the set of possible doses, a bounded interval in $\mathbb{R}$. In this section, we restrict our attention to the case where $n_i = 1$, so the response data are simply $Y_k \in \{Y_{ij} : i = 1, \ldots, N, j = 1\}$ and $n = N$.
The causal analysis is formulated through the use of potential or counterfactual outcomes, that is, the value of the response that would result if a subject were to receive a specified treatment dose not necessarily the same as the dose they received in the study. We denote by \( Y_k(d) \equiv Y_k(D_k = d) \) the potential response random variable resulting from a dose \( D_k = d \) taken in an interval, and write \( y_k(d) \) for the observed version. Potential outcomes adhere to the axiom of consistency: the actual and potential response are equal when the regime in question is the dose actually received, that is, \( y_k(d) = y_k \) if \( d = d_k \).

As with all models for observational data, causal models require certain modelling assumptions to be appropriately specified (Robins, 1997; Robins et al., 2000). Specifically, we make the stable unit treatment value assumption (Rubin, 1978), which states that a subject’s outcome is not influenced by other subjects’ treatment allocation. We further assume weak unconfoundedness of the potential outcomes given the covariates, so that \( Y_k(d) \perp D_k | X_k \) for all \( d \in \mathcal{D} \); see Hirano and Imbens (2004).

Following Imbens (2000) and Hirano and Imbens (2004), we define the observed Generalized Propensity Score (GPS), \( r(d, x) \) for dose \( d \) and covariate values \( x \) by

\[
r(d, x) = f_{D|X}(d|x),
\]

that is, the conditional density function for \( D \) given \( X = x \) evaluated at \( D = d \); \( R = r(D, X) \) is the corresponding random quantity. The GPS is an extension of the propensity score (Rosenbaum and Rubin, 1983) to continuous treatments. In this paper, we regard the construction of this conditional density as a regression problem, and regress \( D \) on \( X \). Note here that, if \( d_1 \neq d_0 \), then we may have \( r(d_1, x_1) \neq r(d_0, x_0) \) even if \( x_1 = x_0 \).

The GPS \( R \), its observed version \( r_k = r(d_k, x_k) \), and potential version \( r_{d,k} = r(d, x_k) \) for \( d \in \mathcal{D} \) form part of the bias removal strategy. As is shown by Hirano and Imbens (2004), the GPS random quantity \( R \) has two properties that render it useful in causal inference problems. First, \( R \) acts as a balancing score - \( D \) and \( X \) are conditionally independent given \( R \); this result implies that within strata of \( R \), the distribution of dose is the same irrespective of the value of the covariate. Secondly, the distribution of the potential response is independent of the covariates, given the treatment and the propensity score - for all sets \( \mathcal{A} \),

\[
\Pr[Y(d) \in \mathcal{A}|D = d, R, X] \equiv \Pr[Y(d) \in \mathcal{A}|D = d, R].
\]

That \( R \) breaks the dependence between \( D \) and \( X \) is the crucial factor that permits causal inference; the conditional independence of \( Y \) and \( X \) given \( D \) and \( R \) permits simplified modelling. We shall see that both features carry over to the longitudinal setting.
2.2. Average Potential Outcomes

To report the causal effect of interest, we thus examine the conditional average causal effect of dose $D$, defined as the difference in expected outcomes for two dose levels $d_0, d_1$ for fixed covariate values $X_k = x$, that is

$$E[Y_k(d_1)|X_k = x] - E[Y_k(d_0)|X_k = x].$$  \hspace{1cm} (2)

The average causal effect is the expectation of this quantity over the distribution of different $X$ values in the study population. The modelling of outcome on dose and the GPS, $R$ (rather than $X$), returns an estimate of

$$E[Y_k(d_1)|R_k = r] - E[Y_k(d_0)|R_k = r]$$  \hspace{1cm} (3)

and its population average. The GPS does not return estimates of the quantity in equation (2), but does yield a bias-removal strategy: we examine the conditional distribution of $Y$ given $D$ and $R$, rather than the conditional distribution given $D$ and $X$, and recover a consistent estimator of the dose-response relationship by averaging appropriately. Specifically, a key quantity of interest in dose-response modelling is the Average Potential Outcome (APO) at dose level $d$, $\mu(d)$, where

$$\mu(d) = E[Y(d)] = E_X[E[Y(d)|X]],$$

which traces the causal dose-response relationship as $d$ varies in $D$. We shall see that equation (3) facilitates consistent estimation of the dose-response relationship, as we may average the conditional expectations over the distribution of $R$ if the balancing property holds, that is, if within-score strata the conditional density value for $D = d$ does not depend on $X$. The adequacy of any proposed propensity score model rests on whether or not balance is achieved, but can be checked by standard exploratory statistical methods.

2.3. An Algorithm for Estimating the APO

The role of the propensity score in estimating the APO is made clear by the identity given in Imbens (2000)

$$\mu(d) = E[Y(d)] = E_X[E[Y(d)|X]] = E_X[E[Y(d)|r(d,X)]] ,$$

where, for fixed dose $d$, the iterated expectation over $Y$ given $X$ and $D$, then $X$, is replaced by an iterated expectation over $Y$ and $R$ and $D$, then $X$, utilizing the fact that for fixed $X$ and $D$, $R$ is completely determined. We outline the estimation procedure, essentially
described in Hirano and Imbens (2004), here.

I. Form the GPS Model: Using a regression approach, construct a predictive model, \( f_{D|X}(d|x, \beta) \), for \( D \) given \( X \). Estimate parameters \( \beta \), from the observed dose and covariate data.

II. Compute the Fitted GPS values: Compute the estimated GPS, \( \hat{r}_k = f_{D|X}(d_k|x_k, \hat{\beta}) \).

III. Form the Observable Model: Using a regression approach, construct a predictive model, \( f_{Y|D,R}(y|d, r, \alpha) \), for \( Y \) given \( D \) and \( R \). Estimate parameters \( \alpha \) using \( \{(y_k, d_k, \hat{r}_k), k = 1, \ldots, N\} \).

IV. Estimate the APO: Estimate the APO at dose level \( d \) by

\[
\hat{\mu}(d) = \hat{E}[Y|D = d] = \frac{1}{n} \sum_{k=1}^{n} E_{Y|D,R}[Y_k(d)|D = d, \hat{r}_k = \hat{r}(d, x_k), \hat{\alpha}]
\]

for \( d \in \mathcal{D} \), where \( \hat{r} \) is evaluated at \( \beta = \hat{\beta} \). Then \( \hat{\mu}(d), d \in \mathcal{D} \) is the GPS-adjusted estimated dose-response function.

Justification for this procedure is given in Hirano and Imbens (2004), and is extended in Section 3.. The two key conditional models \( f_{D|X}(d|x, \beta) \) and \( f_{Y|D,R}(y|d, r, \alpha) \), or the corresponding conditional moments, must be user-specified, but the adequacy of both components can be assessed in a straightforward statistical fashion.

We note that any one-to-one function of the GPS provides the desired balancing property, so that in \( f_{Y|D,R}(y|d, r, \alpha) \), mean response may depend on \( R \), categories defined by discretizing \( R \), or some more general one-to-one function of \( R \). In particular, an alternative approach proposed by Hirano and Imbens (2004) suggests that in practice, the APO may be approximated by estimating the dose-response effect within strata defined by the linear predictor of the treatment density function, and then combining these estimates to form a single, weighted average. This approach is often more straightforward to implement than the above algorithm and often provides an estimate of the dose-response relationship that has little or no biased (see Section 4.).

3. The GPS for Repeated Measures Data

In the case of dose-response estimation from repeated measures or multi-interval data, the potential patterns of confounding are more complex than can be dealt with using a univariate GPS approach. In this section, we formulate a GPS approach suitable for the analysis of
repeated measures response data with interval-dependent dosing. In our by-interval analysis, where interest lies in the estimating the response to dose received within an interval, we will assume that the marginal distribution of the counterfactual response, $Y(d)$, is the same for all intervals, so that we have a single dose-response function $\mu(d) = E[Y(d)]$ to estimate. The method is amenable to more general models, however.

In the repeated measures setting, we no longer wish to ignore the correlation structure in the data, and so we return to the use of notation that makes this explicit. We therefore have that $Y_{ij}$, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$ is the response for individual $i$ in interval $j$; dose and covariate variables are similarly subscripted. Furthermore, we modify the notion of weak unconfoundedness to what we term sequential weak unconfoundedness; dropping the subscript $i$ for convenience, we assume

$$Y_j(d) \perp D_j | X_1, \ldots, X_j.$$ 

That is, at each interval, assignment to dose $d$ is weakly unconfounded with the change in visual acuity during interval $j$ given covariates, previous response, and dose values measured up to the start of the $j$th interval. We denote the history of covariates, response, and previous doses by $\tilde{X}_j = (X_1, \ldots, X_j)^T$.

The GPS procedure must be modified in the case of repeated measures data, in both the construction of the propensity score itself, and in the estimation of the APO. Consider the vector of observed propensity scores, $r$, formed by concatenating the interval-specific observed propensity scores $r_j = f_{D|\tilde{X}_j}(d_j | \tilde{X}_j)$ for each $j$. Given $r$, the conditional expectation in the Observable Model can be estimated using Generalized Least Squares the regression model where for individual $i$,

$$E[Y_i(d) | r_i, d_i] = X_i \alpha$$

where $X_i$ is the design matrix combining vectors $r_i$ and $d_i$ for individual $i$.

We now demonstrate that the repeated measures GPS procedure retains the desired balancing properties of the univariate approach. The theoretical properties of the repeated measures GPS are extended from an adaptation of those in Hirano and Imbens (2004) to cover the multivariate setting. The result for the single interval setting can be recovered from the theorem as a special case.

**Theorem 1 (Weak Unconfoundedness Given the Repeated Measures GPS).** Suppose that assignment to treatment in the $j$th interval is sequentially weak unconfounded given variables $\tilde{X}_j$ that occurred prior to treatment in the current interval (and may include pre-
vious treatment doses). Then, for every dose $d$,

$$f_{D|d(j)}(d|r_{d,j}) = f_{D|R_{d,j},Y_{d,j}}(d|r_{d,j},y_{d,j}).$$

where $Y_{d,j} = Y_j(d)$ and $y_{d,j} = y_j(d)$.

**Proof:** Sequential weak unconfoundedness implies that for all $d \in D$, $Y_j(d) \perp D_j|\tilde{X}_j$, that is, for each dose $d$, $Y_j$ and $D_j$ are conditionally independent given $\tilde{X}_j$. Consider the random quantities, $R_j = r(D_j, \tilde{X}_j)$, where $r(d, \tilde{x}_j) = f_{D|R}(d|\tilde{x}_j)$, and $R_{d,j} = r(d, \tilde{X}_j)$, defined for fixed $d$. Denoting by $f$ the density function for the relevant random variables, we have

$$f_{D|R_{d,j}}(d|r_{d,j}) = \int_{X_{d,j}} f_{D|R_{d,j}}(d, \tilde{x}_j|r_{d,j}) d\tilde{x}_j = \int_{X_{d,j}} f_{D|R_{d,j}}(d|\tilde{x}_j, r_{d,j}) f_{\tilde{x}_j|R_{d,j}}(\tilde{x}_j|r_{d,j}) d\tilde{x}_j$$

where $X_{d,j} \subset X_j$ is the set of solutions $\tilde{x}_j$ of the equation $r_{d,j} = r(d, \tilde{x}_j)$. Now, on $X_{d,j}$,

$$f_{D|R_{d,j}}(d|\tilde{x}_j, r_{d,j}) = f_{D|\tilde{x}_j}(d|\tilde{x}_j) = r_{d,j},$$

as for fixed $d$ and $\tilde{x}_j$ on the contour, $r_{d,j}$ is completely defined. Thus

$$f_{D|R_{d,j}}(d|r_{d,j}) = \int_{X_{d,j}} r_{d,j} f_{\tilde{x}_j|R_{d,j}}(\tilde{x}_j|r_{d,j}) d\tilde{x}_j = r_{d,j} \int_{X_{d,j}} f_{\tilde{x}_j|R_{d,j}}(\tilde{x}_j|r_{d,j}) d\tilde{x}_j = r_{d,j} = f_{D|\tilde{x}_j}(d|\tilde{x}_j)$$

for an archetypal $\tilde{x}_j$ on $X_{d,j}$. Similarly, by weak unconfoundedness,

$$f_{D|R_{d,j},Y_{d,j}}(d|r_{d,j}, y_{d,j}) = \int_{X_{d,j}} f_{D|R_{d,j},Y_{d,j}}(d|\tilde{x}_j, r_{d,j}, y_{d,j}) f_{\tilde{x}_j|R_{d,j},Y_{d,j}}(\tilde{x}_j|r_{d,j}, y_{d,j}) d\tilde{x}_j$$

$$= \int_{X_{d,j}} f_{D|\tilde{x}_j}(d|\tilde{x}_j) f_{\tilde{x}_j|R_{d,j},Y_{d,j}}(\tilde{x}_j|r_{d,j}, y_{d,j}) d\tilde{x}_j$$

$$= \int_{X_{d,j}} r_{d,j} f_{\tilde{x}_j|R_{d,j},Y_{d,j}}(\tilde{x}_j|r_{d,j}, y_{d,j}) d\tilde{x}_j = r_{d,j} = f_{D|\tilde{x}_j}(d|\tilde{x}_j)$$

for an archetypal $\tilde{x}_j$ on $X_{d,j}$. Thus, for all $d$, $f_{D|R_{d,j}}(d|r_{d,j}) = f_{D|R_{d,j},Y_{d,j}}(d|r_{d,j}, y_{d,j})$ and we have weak unconfoundedness given $R_{d,j} = r(d, \tilde{X}_j)$.

**Theorem 2 (Bias Removal of the Repeated Measures GPS Procedure).** Suppose that $\mu(d) = E[Y(d)]$ be the marginal mean of interest. For interval $j$, consider the mean

$$\beta(d, r_{d,j}) = E[Y_j(d)|D = d, R_{d,j} = r_{d,j}]$$

that conditions on the GPS. The Average Potential Outcome, obtained by averaging $\beta(d, r_{d,j})$ over the observed distribution of the covariates $\tilde{X}_j$, is an unbiased estimator of the dose-response function $\mu(d)$.
Proof: By conditional probability and Theorem 1 above,

\[
f_{Y_{d,j}|D,R_{d,j}}(y_{d,j}|d,r_{d,j}) = \frac{f_{Y_{d,j}|R_{d,j}}(y_{d,j}|r_{d,j}) f_D|R_{d,j}(d|r_{d,j})}{f_D|R_{d,j}(d|r_{d,j})} = \frac{f_{Y_{d,j}|R_{d,j}}(y_{d,j}|r_{d,j}) f_D|R_{d,j}(d|r_{d,j})}{f_D|R_{d,j}(d|r_{d,j})} = f_{Y_{d,j}|R_{d,j}}(y_{d,j}|r_{d,j})
\]

so that \( E[Y_j(d)|D = d, R_{d,j} = r_{d,j}] = E[Y_j(d)|R_{d,j} = r_{d,j}] \). However,

\[
E[Y_j(d)|D = d, R_{d,j} = r_{d,j}] = E[Y_j(d)|D = d, r(D, \tilde{X}_j) = r_{d,j}] = E[Y_j(d)|D = d, r(d, \tilde{X}_j) = r_{d,j}] = E[Y_j(d)|D = d, R_{d,j} = r_{d,j}] = \beta(d, r_{d,j})
\]

using the first result. Thus, by iterated expectation, noting that \( E[Y(d)] \equiv E[Y_j(d)] \),

\[
\mu(d) \equiv E[Y_j(d)] = E_{R_{d,j}}[E[Y_j(d)|R_{d,j} = r_{d,j}]] = E_{R_{d,j}}[\beta(d, R_{d,j})] \equiv E_{\tilde{X}_j}[\beta(d, r(d, \tilde{X}_j))].
\]

Corollary: By Theorem 2, applying the bias removal result sequentially to each interval, we obtain an unbiased estimator of \( \mu(d) \) after pooling results over all intervals, by taking the expectation in turn over \( \tilde{X}_1, \tilde{X}_2, \ldots \).

NOTE: A univariate GPS analysis that does not construct \( R \) by conditioning on \( \tilde{X}_j = \tilde{x}_j \) for each \( j \) does not necessarily achieve bias removal.

4. Simulation Studies and Examples

4.1. Simulation I: Nonlinear, nonadditive treatment effect

Here, we extend the artificial example of Hirano and Imbens (2004) to a two-interval setting. The causal structure of the model is depicted in Figure 1.

Data generation: Suppose that at the first and second interval, we have

\[
Y_1(d)|X_{11}, X_{12} \sim \mathcal{N}(d + (X_{11} + X_{12}) \exp[-d(X_{11} + X_{12})], 1)
\]

\[
Y_2(d)|X_{21}, X_{12} \sim \mathcal{N}(d + (X_{21} + X_{12}) \exp[-d(X_{21} + X_{12})], 1)
\]

Suppose that the marginal distributions of each of \( X_{11}, X_{12}, \) and \( X_{21} \) are all unit exponential, and the marginal mean of the response in either interval is identical. As in Hirano and Imbens
(2004), the APO can be obtained by integrating out the covariates analytically, yielding
\[
\mu(d) = d + \frac{2}{(1+d)^3}.
\]
A multivariate GPS analysis, will involve the concatenated GPS vector \( R^M = (R_1, R_2)^T \) where \( R_1 = (X_{11} + X_{12}) \exp[-D(X_{11} + X_{12})] \) and \( R_2 = (X_{21} + X_{12}) \exp[-D(X_{21} + X_{12})] \), which consists of correctly-specified models. A univariate, or cross-sectional, GPS analysis might fail to include information from the previous interval and hence the GPS used would be \( R^U = (R_1, R_2^*)^T \) where \( R_1 \) is as before, but \( R_2^* = X_{21} \exp[-DX_{21}] \).

**Analysis and Results:** We generated 1,000 datasets of size 250, 500, 100, and 10,000. The mean and median APO/dose-response curves using the MGPS were exactly correct, while the UPGS analysis (see Figure 2 for results with \( n = 250 \)) was clearly biased. The general shape of the UGPS APO was correct, however the curve fell outside of the confidence bands of the (unbiased) MGPS over part of the range of doses even with samples as small as 250.

**4.2. Simulation II: Nonlinear, nonadditive confounding**

We now demonstrate the need for repeated measures GPS model by considering a plausible potential confounding mechanism. Suppose that data are distributed according to the causal structure defined in Figure 3. That is, treatment dose received in an interval is caused by (or is a descendant of) the response level \( Y \) measured prior to the dose being taken and of dose taken in the previous interval (when not in the first treatment interval).

We generalize the univariate setting used by others (Rubin and Thomas, 2000; Imai and Van Dyk, 2004), in which the exponential function is used to specify a mean model which the effects of the confounding variables on \( Y \) in each interval are nonlinear and nonadditive.

**Data Generation:** At each interval \( j, j = 1, 2 \), variables \( X_{j1}, X_{j2}, X_{j3}, X_{j4} \) are measured. The variables \( X_{j1}, X_3, \) and \( X_{j4} \) are Normally distributed, while \( X_{j2} \) is a Bernoulli random
variable. At each interval, treatment dose $D_j$ and response $Y_j$ are distributed as follows:

$$D_1 \sim \frac{1}{100}(400.9 - 0.8X_{11}^2 + 1.6X_{12} - 0.65X_{13}) + \epsilon_1$$
$$Y_1 \sim -18.7 + 2.3D_1 + \exp(-0.8X_{11} + 2.3X_{12} - 1.1X_{13}) \times \mathbb{I}[\exp(-0.8X_{11} + 2.3X_{12} - 1.1X_{13}) < 50] - 4.2X_{14} + 10\xi_1$$

$$D_2 \sim \frac{1}{100}(400.9 - 0.8X_{21}^2 + 1.6X_{22} - 0.65X_{23}) + Y_1 + \epsilon_2$$
$$Y_2 \sim -18.7 + 2.3D_2 + \exp(-0.8X_{21} + 2.3X_{22} - 1.1X_{23} - 0.2Y_1) \times \mathbb{I}[\exp(-0.8X_{21} + 2.3X_{22} - 1.1X_{23} - 0.2Y_1) < 50] - 4.2X_{24} + 10\xi_2$$

where $\epsilon_1, \xi_1, \epsilon_2, \xi_2 \sim \mathcal{N}(0, 1)$ are mutually independent. The coefficients in these models were chosen arbitrarily, with indicator functions included to prevent extremely large responses.

Let $Y = (Y_1, Y_2)^T$, $D = (D_1, D_2)^T$, and $X_l = (X_{1l}, X_{2l})^T$, for $l = 1, 2, 3, 4$. We may distinguish these covariates as variables which confound, $X^c = (X_{11}, X_{12}, X_{13})$, and the variable which only predicts response (not treatment), $X^p = X_{14}$.

**Analyses:** We apply the univariate and multivariate GPS functions assuming a Normal distribution of dose, and in each of analyses – a non-causal regression model, a univariate GPS approach, and a multivariate GPS approach – the mean model for response in an interval will be (incorrectly) assumed to be linear. A fourth approach will also be considered: the approximation to the APO suggested by Imai and Van Dyk (2004) based on the multivariate GPS model. Specifically, we perform the following five analyses:

1. **Linear model:** $E[Y]$ is fit as a linear function of $D$, $X^c$, and $X^p$.

2. **Univariate GPS** (UGPS): $E[D]$ is fit as a linear function of $X^c$ and used to construct the GPS $R^U$; $E[Y]$ is fit as a linear function of $D$, $R^U$, and $X^p$.

3. **Repeated measures**, or **multivariate, GPS** (MGPS Linear): $E[D_1]$ is fit as a linear function of $X_{11}, X_{12}, X_{13}$, and $X_{14}$ while $E[D_2]$ is fit as a linear function of $X_{21}, X_{22}, X_{23}, X_{24}$, and $Y_1$. These models are used to construct the GPS $R^M$, consisting of the concatenated Generalized Propensity Scores at each interval. $E[Y]$ is fit as a linear function of $D$, $R^M$, and $X^p$.

4. **Repeated measures GPS** (MGPS Quintiles): The GPS $R^M$ is estimated as in (3) above, however the GPS is now used to create dummy variables for membership in each quintiles of the observed GPS values. $E[Y]$ is fit as a linear function of $D$, the variables for quintile membership, and $X^p$. 

11
5. **Repeated measures GPS** (Linear Predictor Quintiles): The GPS $R^M$ is estimated as in (3) above, however the *linear predictor* of the GPS is used to create dummy variables for membership in each quintiles of the observed GPS values. $E[Y]$ is fit as a linear function of $D$, the variables for quintile membership, and $X^p$.

The univariate GPS approach, which controls for the confounding attributable to the covariates but ignores potential confounding by previous responses, will provide adequate balance for estimating the cross-sectional effect of dose at the first interval but will not provide the required balance in the second interval, where confounding of the cross-sectional relationship between $D_2$ and $Y_2$ by $Y_1$ is observed. The repeated measures GPS, when constructed so as to correctly model the causal structure of the data, has the same balancing properties as the univariate GPS, in that within strata of $R$, the distribution of $D$ does not depend on $X$. However, both GPS analyses rely on the incorrectly specified mean models for both dose and response. The (non-causal) linear is also mis-specified. One thousand data sets were generated for sample sizes 250, 500, 1000, and 10,000.

**Results:** In all sample sizes, the repeated measures GPS approach using the GPS as a linear term in the response model produced unbiased estimates of the true dose effect: at the smallest sample size ($n = 250$), bias was less than 0.91%, and dropped to 0.08% with $n = 10,000$. The repeated measures GPS approach using the GPS to construct dummy variables of quintile membership exhibited bias of about 3-5%, followed distantly by the univariate GPS approach (18%) and the linear model (20-21%). The repeated measures GPS approach using the linear predictor of the GPS to construct dummy variables of quintile membership were virtually identical to those using the quintiles of the GPS (results not shown); this is to be expected when the GPS model provides a reasonable fit to the doses.

It is interesting to note that the linear model, the univariate GPS, and the repeated measures GPS (linear GPS) approach all produce unbiased estimates of the effect of the predictor of response, $X_4$, which does not predict treatment. The bias was less than 0.5% with each of the aforementioned approaches. The variability of the estimate from the linear model was smallest, though only slightly less than the variability of the repeated measures GPS (linear GPS). The repeated measures GPS approach using quintile membership produced highly biased estimates of the effect of $X_4$ on response.

The performance of the repeated measures GPS approach using quantiles of the GPS as dummy variables may be improved by taking a greater number of quantiles. Using deciles rather than quantiles provides dose-effect estimates that exhibit very little bias (below 3%). The variability of the dose-effect estimate using this approach is, however, larger than that found by including the GPS as a continuous variable.
4.3. Example I: Riesby Depression Study

Reisby et al. (1977) treated 66 hospitalized depressed patients for four weeks with 225 mg/day imipramine; patients were classified as having endogenous or non-endogenous depression. Antidepressive effect was evaluated on the basis of the post-treatment rating scores. We analyze the effect of imipramine concentration measured weekly on the change in depression, as assessed by the Hamilton's Depression Rating Scale (HDRS).

**Analysis:** The concentration of imipramine in the blood plasma is approximately Normally distributed on the log scale. A univariate GPS analysis was performed using log-imipramine concentration as a linear function of patient sex and depression type (endogenous or not) and an interaction between these variables, and then estimating the APO using a linear regression of the change in HDRS on log-imipramine concentration and the univariate GPS.

An MGPS analysis was also performed under the assumption of first-order Markov property for the log-imipramine concentrations. That is, log-imipramine concentrations within (and between) patients were assumed to be conditionally independent, given the HDRS score at the start of the week, patient sex, and depression type. The repeated measures APO was computed using a linear mixed effect model, regressing the change in HRS on log-imipramine concentration and the repeated measures GPS and allowing for random intercepts.

**Results:** Both the univariate and repeated measures GPS analyses indicate that depression decreases with increasing blood concentrations of imipramine on the range of 0.47 - 1.75 units (or 1.60-5.75 log-units), as observed in Figure 4. The dose-response relationship shows an approximately linear decrease in symptoms across the range of log concentrations observed. The univariate and repeated measures analyses produce nearly identical dose-response curves. This is not surprising, as exploratory plots revealed very little correlation between HDRS at the start of the week and either the log-imipramine concentration during the week or the change in HDRS during the week.

While few covariates were available for this analysis, raising the possibility of unmeasured confounding, it was plausible a priori that the Riesby data follow a structure akin to that in Figure 3. That is, the change in depression score may be caused by the current level of depression and as well as the drug received.

One might speculate whether the concentration of imipramine in the previous interval also confounds the relationship between change in HDRS and the most recent concentration of imipramine. Exploratory plots reveal a strong correlation between drug concentration from one interval to the next, but virtually none between change in HDRS and the concentration.
of imipramine in the previous interval. Recent work in the binary-treatment propensity score literature suggests that including variables which predict treatment but are unrelated to response can create bias (Austin et al., 2007; Brookhart et al., 2006), therefore imipramine concentration in the previous interval was not included in the either of the GPS models.

4.4. Example II: MOTAS Amblyopia Study

Amblyopia is the most common childhood vision disorder, and is characterized by reduced visual function in one eye. A standard treatment for the condition is occlusion therapy, that is, patching of the functioning fellow eye. The apparent beneficial effect of occlusion therapy has never been well quantified, partly due to difficulty in the accurate measurement of the occlusion dose. The Monitored Occlusion Treatment of Amblyopia Study (MOTAS) (Stewart et al., 2004) was the first clinical study aimed at quantifying the dose-response relationship of occlusion, facilitated by the use of an electronic occlusion dose monitor.

The MOTAS design and a full description of the study base have been published previously (Stewart et al., 2002, 2004). At study entry, all children who required spectacles entered the refractive adaptation phase; the remainder entered the occlusion phase directly. Children still considered amblyopic after refractive adaption began occlusion and were prescribed six hours of occlusion daily. Visual acuity was measured on the logarithm of Minimum Angle of Resolution (logMAR) scale; improvement is indicated by a decrease in logMAR. Visual function and monitored occlusion dose were recorded at approximately two-week intervals until acuity ceased to improve. A total of 116 children were enrolled in MOTAS; we analyze data of the 68 who took part in the occlusion phase who, although prescribed six hours of occlusion daily, received varying occlusion doses because of incomplete concordance.

Notation for MOTAS: For child \(i\), the response, \(Y_{ij}\), is the change visual acuity during interval \(j\) for patient \(i\), and \(D_{ij}\) be the random occlusion dose (in hours) received in interval \(j\). Let \(A_{ij}\) be the age in months at the start of interval \(j\) and \(L_{ij}\) and \(P_i\) denote the visual acuity at the start of interval \(j\) and at the start of the phase, respectively.

In the study, dose is a continuous variable, but 60 out of 404 (about 15%) of intervals in the occlusion phase had a zero dose. The GPS model \(f_{D|X}(d|x, \beta)\) must acknowledge the mixture nature of the dose distribution, so we assume that, given \(X = x\),

\[
D = \frac{x}{\pi(x, \gamma)}[d = 0] + (1 - \pi(x, \gamma))[d \neq 0]D_+\tag{4}
\]

where \(D_+\) is a strictly positive random variable whose distribution depends on \(X = x\) and \(\beta\), and \(0 < \pi(x, \gamma) < 1\) is a mixing weight. Estimation in this model is straightforward.
when a parametric distribution is used for \( D_+ \), and such regression model that induces a balancing property can be used. To estimate \( \gamma \), we fit a logistic regression model to the binary \( (D = 0)/D > 0 \) dose data.

Recent work has shown that optimal (binary treatment) propensity scores include all confounding variables and variables that predict outcome, while variables that are purely predictors of treatment should not be included in the model (Brookhart et al., 2006; Austin et al., 2007). The following covariates therefore included: visual acuity at start of interval, \( L \), interval number, length of interval (in days), and amblyopic type (anisometropic, strabismic, mixed). These covariates were used to predict both the probability of having any occlusion at all \( (D/D > 0) \) in a logistic model and the probability of receiving a particular dose (greater than zero) of occlusion in a Weibull model. The GPS used was

\[
\hat{r}(d, x) = \hat{\pi}(x, \hat{\gamma}) \mathbb{I}[d = 0] + (1 - \hat{\pi}(x, \hat{\gamma})) \mathbb{I}[d \neq 0] f(d|x, \hat{\phi}, \hat{\beta})
\]

where \( f(d|x, \hat{\phi}, \hat{\beta}) \) is a Weibull density with shape \( \hat{\phi} \) and scale \( \exp\{x^T \beta\} \). For the GPS to act as a balancing score, the distribution of \( D \) should not depend on \( \hat{X} \) within strata of \( \hat{r} \). The balancing property appears to have been approximately achieved.

As response in the MOTAS is the vector of changes in visual acuity, there is little observed serial correlation in the data, so we ignore the repeated measures nature of the data. The observable model for change in visual acuity, \( Y \), is modelled via the expectation

\[
E_{Y|D,R}(Y|D = d, R = r, \alpha) = \alpha_0 + \mathbb{I}[r < 0.1]\left(\alpha_1 + \alpha_2 d + \alpha_3 r + \alpha_4 d r\right).
\]

This model can be readily extended to a more flexible or piecewise constant partition model. However, here, the addition of higher order terms led to only minimal changes in the inferences made. Also, when using a mixture distribution such as (4) for the GPS, it may be that the \( \hat{r} \) values for one component differ substantially from those of the other, so that there are no data in a portion of the space consisting of all \((d, r)\)-pairs for \( d \in D, r \in \text{range}(\hat{R}) \) where \( \text{range}(\hat{R}) \) denotes the range of estimated generalized propensity scores. We account for this explicitly in the model; rather than fitting a model that assumes that the relationship between response and dose and the GPS is the same function in regions of the plane where \((d, r)\) pairs were observed, and in regions where no data was observed. Using the model in equation (5), we obtain least-squares estimates (SE) \( \hat{\alpha}_0 = -0.018 \ (0.008), \hat{\alpha}_1 = 0.009 \ (0.041), \hat{\alpha}_3 = -3.19-04 \ (3.24e-04), \hat{\alpha}_3 = -0.083 \ (4.042), \) and \( \hat{\alpha}_4 = -0.101 \ (0.080) \), respectively.

**Results of the UGPS Analysis:** A plot of the dose-response curve is presented in Figure 5(a). This plot indicates that the association between dose and visual acuity, when confound-
ing between dose and the covariates is adjusted for using the GPS approach, is appreciable; the average potential effect on change in visual acuity measurement $Y$ is significantly negative (corresponding to vision improvement) over the entire range of positive doses considered. A numerical summary is given in Table 2.

A plot of the average (over the covariate distribution) dose effect estimated by a non-causal mixed effects model analysis is presented in Figure 5(a) for comparison. The intercept of the causal APO is nearer to zero than the regression ADE, in agreement with ophthalmological belief that visual acuity will not improve spontaneously in the absence of occlusion. Also, the APO suggests a plateau, or a saturation of the effect of occlusion in an interval at about 80 hours. This indicates that children may not exhibit a clinically meaningful improvement in visual acuity with more than, on average, six hours of occlusion per day over a two-week period. This conclusion is entirely reasonable, as physical changes to the amblyopic eye that can occur in a fixed time period are likely limited by biological processes.

**Applying the MGPS to the MOTAS Data:** There is evidence in the data to suggest considerable correlation between dose of occlusion in an interval and that received in the previous interval (estimated correlation, unadjusted for within-person clustering, is 0.45). We adopt a first-order Markov model, and model dose in each interval as a function of the dose received in the previous interval. We adopt a GPS model similar to the model in equation (4) modified for the longitudinal setting; for interval $j$,

$$D_j = \pi(\bar{x}_j, d_{j-1}, \gamma)[d_j = 0] + (1 - \pi(\bar{x}_j, d_{j-1}, \gamma))[d_j \neq 0]D_{j+}$$

with, again, a Weibull model dependent on $\bar{X}_j, D_{j-1}$ for the continuous part of the distribution and use logistic regression model to the binary ($D = 0/ D > 0$) dose data.

The following covariates were included: occlusion dose in the previous interval, visual acuity at start of interval, $L$, interval number, length of interval (in days), and amblyopic type. Coefficient estimates (SE) are displayed in Table 3. A graphical check of whether the balancing property was achieved raised no concerns. The observable model in equation (5) was again adopted. Parameters were estimated using a linear mixed effects model with a random intercept to account for any correlation that may exist in the response (though little is expected, as the response measures the change in visual acuity). Using the model in equation (5), we obtain estimates (SE) $\hat{\alpha}_0 = -0.020$ (0.008), $\hat{\alpha}_1 = -0.002$ (0.040), $\hat{\alpha}_2 = -3.05e-4$ (3.19e-4), $\hat{\alpha}_3 = 0.218$ (4.103), and $\hat{\alpha}_4 = -0.062$ (0.085), respectively.

**Results of the MGPS Analysis:** A plot of the dose-response curve is presented in Figure 5(b); a numerical summary of the APO based on a repeated measures GPS analysis is
provided in Table 4. The average change in visual acuity due to occlusion predicted by the repeated measures GPS modelling is very similar to that predicted by the univariate GPS model. The similarity of results is due to the fact that the univariate GPS provided the required balancing property, and so little additional reduction in bias was anticipated. It appears that in MOTAS, dose in the previous interval is highly correlated with current dose of occlusion, but not strongly associated with current change in visual acuity.

5. Discussion

In a longitudinal study of dose-response, full compliance is the exception rather than the expected. To estimate the dose-response relationship with confidence, modelling potentially confounding relationships flexibly is key. Generalized Propensity Scores are under-used in the cross-sectional (single time-point) setting, and yet provide a tractable and flexible option of analysis, and can be adapted to any number of complex dosing strategies. We have extended the GPS methodology to the repeated measures setting to cope with situations where treatment doses received in different intervals are correlated and response may depend on doses in current and earlier intervals.

In applying the repeated measures GPS approach, one must take care over the issue of the intercept of the causal dose-response curve. For example, if in the Riesby example response was taken to be depression rating score rather than the change in depression score, the dose-response curve estimated over all intervals similarly would be averaging a series of parallel dose response curves with different intercepts (since, from interval to interval, the average depression would show improvement due to dose in the previous interval).

Whether analyzing data from a cross-sectional study or in repeated measures setting, it is essential to consider a response model that is flexible in its dependence on the GPS or to use the approximate method of estimating the APO. The GPS often incorporates covariate information into the mean model for response in a highly nonlinear fashion. An incorrectly-specified mean model may not provide a good estimate of the true dose-response curve, even if the GPS has successfully achieved the desired balancing of confounding variables.

ACKNOWLEDGMENT: Both authors acknowledge funding through the Natural Sciences and Engineering Research Council of Canada. MOTAS was supported by The Guide Dogs for the Blind Association, UK.


Table 1: Simulation II results: Coefficient estimates (standard errors) for the parameters from a linear model, a univariate GPS analysis (UGPS), and multivariate GPS analyses (MGPS) where the GPS is included as a linear term in the response model or quintiles of the GPS are included as factor level variables in a linear model. The true coefficient of $D$ is 2.3 in all cases.

<table>
<thead>
<tr>
<th>n</th>
<th>Model:</th>
<th>Linear model</th>
<th>UGPS</th>
<th>MGPS (Linear)</th>
<th>MGPS (Quintiles)</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>Intercept</td>
<td>-11.60 (3.83)</td>
<td>-36.15 (29.15)</td>
<td>-6.17 (8.66)</td>
<td>-13.68 (5.21)</td>
</tr>
<tr>
<td></td>
<td>$D$</td>
<td>1.810 (0.559)</td>
<td>1.889 (0.565)</td>
<td>2.321 (1.093)</td>
<td>2.210 (1.279)</td>
</tr>
<tr>
<td></td>
<td>$X_1$</td>
<td>-1.262 (1.018)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$X_2$</td>
<td>3.710 (2.082)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$X_3$</td>
<td>-1.942 (2.157)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$X_4$</td>
<td>-4.187 (0.710)</td>
<td>-4.189 (0.700)</td>
<td>-4.187 (0.698)</td>
<td>-2.680 (14.142)</td>
</tr>
<tr>
<td></td>
<td>GPS</td>
<td>5.625 (7.572)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>Intercept</td>
<td>-11.58 (2.34)</td>
<td>-46.90 (26.10)</td>
<td>-6.76 (6.51)</td>
<td>-13.46 (3.57)</td>
</tr>
<tr>
<td></td>
<td>$D$</td>
<td>1.817 (0.405)</td>
<td>1.898 (0.404)</td>
<td>2.297 (0.554)</td>
<td>2.221 (0.867)</td>
</tr>
<tr>
<td></td>
<td>$X_1$</td>
<td>-1.276 (0.506)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$X_2$</td>
<td>3.730 (1.186)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$X_3$</td>
<td>-1.884 (0.803)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$X_4$</td>
<td>-4.199 (0.469)</td>
<td>-4.197 (0.485)</td>
<td>-4.195 (0.485)</td>
<td>-3.049 (5.365)</td>
</tr>
<tr>
<td></td>
<td>GPS</td>
<td>8.405 (6.789)</td>
<td></td>
<td></td>
<td>-2.380 (1.820)</td>
</tr>
<tr>
<td>1000</td>
<td>Intercept</td>
<td>-11.57 (1.88)</td>
<td>-54.34 (22.56)</td>
<td>-7.65 (4.76)</td>
<td>-13.59 (2.49)</td>
</tr>
<tr>
<td></td>
<td>$D$</td>
<td>1.816 (0.306)</td>
<td>1.895 (0.306)</td>
<td>2.286 (0.383)</td>
<td>2.178 (0.620)</td>
</tr>
<tr>
<td></td>
<td>$X_1$</td>
<td>-1.275 (0.392)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$X_2$</td>
<td>3.721 (0.932)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$X_3$</td>
<td>-1.998 (0.673)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$X_4$</td>
<td>-4.192 (0.397)</td>
<td>-4.190 (0.408)</td>
<td>-4.190 (0.409)</td>
<td>-3.040 (3.494)</td>
</tr>
<tr>
<td></td>
<td>GPS</td>
<td>10.339 (5.865)</td>
<td></td>
<td></td>
<td>-2.138 (1.347)</td>
</tr>
<tr>
<td>10,000</td>
<td>Intercept</td>
<td>-11.59 (0.55)</td>
<td>-63.24 (8.45)</td>
<td>-9.79 (2.26)</td>
<td>-13.40 (0.80)</td>
</tr>
<tr>
<td></td>
<td>$D$</td>
<td>1.836 (0.108)</td>
<td>1.916 (0.108)</td>
<td>2.298 (0.131)</td>
<td>2.189 (0.407)</td>
</tr>
<tr>
<td></td>
<td>$X_1$</td>
<td>-1.297 (0.138)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$X_2$</td>
<td>3.774 (0.321)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$X_3$</td>
<td>-1.943 (0.239)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$X_4$</td>
<td>-4.205 (0.122)</td>
<td>-4.206 (0.125)</td>
<td>-4.207 (0.126)</td>
<td>-3.305 (1.128)</td>
</tr>
<tr>
<td></td>
<td>GPS</td>
<td>12.624 (0.125)</td>
<td></td>
<td>-1.595 (0.622)</td>
<td></td>
</tr>
</tbody>
</table>
Table 2: MOTAS Data: Summaries of the APO (on the logMAR scale) from a univariate GPS model analysis for changing dose amount per interval: 5000 bootstrap samples.

<table>
<thead>
<tr>
<th>Quantile</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>75</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequentist</td>
<td>-0.036</td>
<td>-0.051</td>
<td>-0.077</td>
<td>-0.091</td>
<td>-0.104</td>
</tr>
<tr>
<td></td>
<td>-0.025</td>
<td>-0.039</td>
<td>-0.063</td>
<td>-0.076</td>
<td>-0.085</td>
</tr>
<tr>
<td></td>
<td>-0.020</td>
<td>-0.033</td>
<td>-0.056</td>
<td>-0.069</td>
<td>-0.076</td>
</tr>
<tr>
<td></td>
<td>-0.014</td>
<td>-0.028</td>
<td>-0.049</td>
<td>-0.062</td>
<td>-0.068</td>
</tr>
<tr>
<td></td>
<td>0.003</td>
<td>-0.017</td>
<td>-0.037</td>
<td>-0.050</td>
<td>-0.053</td>
</tr>
</tbody>
</table>

Table 3: MOTAS Data: Estimates and standard errors for the parameters from the repeated measures (first-order Markov) GPS model: the model comprised a logistic regression for \( D = 0 \) versus \( D > 0 \) and a Weibull model for positive dose.

<table>
<thead>
<tr>
<th>Model: Any dose</th>
<th>Continuous dose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term</td>
<td>Est. (SE)</td>
</tr>
<tr>
<td>Intercept</td>
<td>-2.527 (0.514)</td>
</tr>
<tr>
<td>Previous dose</td>
<td>-0.014 (0.002)</td>
</tr>
<tr>
<td>Visual acuity</td>
<td>0.247 (0.470)</td>
</tr>
<tr>
<td>Interval number</td>
<td>0.220 (0.048)</td>
</tr>
<tr>
<td>Interval length</td>
<td>0.013 (0.007)</td>
</tr>
<tr>
<td>Type: mixed</td>
<td>-1.812 (0.447)</td>
</tr>
<tr>
<td>Type: strabismic</td>
<td>-0.239 (0.445)</td>
</tr>
</tbody>
</table>

Table 4: MOTAS Data: Summaries of the APO (on the logMAR scale) from a repeated measures GPS model analysis for changing dose amount per interval: 5000 bootstrapsamples.

<table>
<thead>
<tr>
<th>Quantile</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>75</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequentist</td>
<td>-0.040</td>
<td>-0.054</td>
<td>-0.087</td>
<td>-0.105</td>
<td>-0.116</td>
</tr>
<tr>
<td></td>
<td>-0.025</td>
<td>-0.042</td>
<td>-0.072</td>
<td>-0.089</td>
<td>-0.097</td>
</tr>
<tr>
<td></td>
<td>-0.018</td>
<td>-0.035</td>
<td>-0.064</td>
<td>-0.080</td>
<td>-0.088</td>
</tr>
<tr>
<td></td>
<td>-0.011</td>
<td>-0.029</td>
<td>-0.056</td>
<td>-0.072</td>
<td>-0.079</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>-0.018</td>
<td>-0.042</td>
<td>-0.056</td>
<td>-0.062</td>
</tr>
</tbody>
</table>
Figure 1: Simulation Study I: Directed Acyclic Graph of the longitudinal structure of simulated data for a nonlinear, nonadditive treatment effect generative model.
Figure 2: Simulation Study I: Pointwise median APO from univariate GPS (UGPS) and repeated measures GPS analyses (MGPS) with point-wise 95% credible intervals the MGPS APO for a nonlinear, nonadditive treatment effect generative model.
Figure 3: Simulation Study II: Directed Acyclic Graph of the longitudinal structure of simulated data for a generative model with nonlinear, nonadditive confounding:
Figure 4: Riesby Data: The estimated average potential change in Hamilton Depression Rating Scale (HDRS) for log-imipramine blood concentrations in the range of 1.60 to 5.75 units with point-wise 95% credible intervals for univariate GPS (UGPS) and repeated measures GPS analyses (MGPS). Observed dose values indicated along the horizontal axis.
Figure 5: MOTAS Data: The estimated average potential change in visual acuity (APO) for doses in the range of 1 to 100 hours per interval with point-wise 95% credible interval. In (a), the univariate GPS (UGPS) APO of Section 2. with the average dose effect (ADE) from a non-causal regression model included for comparison. In (b), the repeated measures GPS (MGPS) APO of Section 3. is plotted, with the UGPS APO included in gray for comparison.