Quantification of the Dose-Response Relationship for a Continuous Treatment in the Presence of Confounding or Informative Non-Compliance

E. E. M. Moodie

Department of Epidemiology, Biostatistics and Occupational Health, McGill University

and

D. A. Stephens*

Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Street West, Montreal, Quebec, Canada, H3A 2K6

September 6, 2006

Summary. In a longitudinal study of dose-response, the presence of confounding or non-compliance compromises the estimation of the true effect of a treatment. Flexibility in modelling confounding is essential in order to capture the treatment effect when the causal model is not fully understood, so that observed treatment effect is not due to the imposition of a rigid model for the relationship between response, treatment and other covariates. A semiparametric additive linear mixed (SPALM) model (Ruppert et al., 2003) provides a tractable and flexible approach to modelling the influence

* *email:* d.stephens@math.mcgill.ca

of potentially confounding variables. However, this approach does not on its own remove the bias introduced due to patient-selected treatment level, that is, it does not permit the estimation of the causal effect of dose. Using an approach based on the Generalized Propensity Score (GPS) (Hirano and Imbens, 2004), a generalization of the classical, binary treatment propensity score, it is possible to construct instrumental variables that provide a more meaningful (and less biased) estimation procedure for the true effect of dose. In this paper, we present Bayesian versions of the SPALM model and of the GPS where the propensity score relies on a novel formulation of the treatment density. The use of Bayesian methods are readily implementable and allow cohesive propagation of uncertainty in the models. The methodology is applied to the Monitored Occlusion Treatment of Amblyopia Study (MOTAS) which investigated the dose-response relationship between occlusion and improvement in visual acuity. This analysis quantifies the beneficial effect of occlusion for the first time.

KEY WORDS: Semi-parametric Modelling, Confounding, Non-compliance, Generalized Propensity Score, Bayesian Analysis

1. Introduction

In observational studies of the efficacy of a treatment, if the treatment dose level is influenced by subject-specific covariates, then there is the potential for bias in the estimation of the treatment effect. Furthermore, many randomized trials, particularly those where treatment is given over time, 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 non-random nature of non-compliance. The aim of this paper is to provide a framework for examining the effect of treatment given over time in the presence of "informative" non-compliance due to the influence of measured covariates.

In this paper, we analyze the data from a recent observational study of the treatment of childhood amblyopia - a common ophthalmological condition, where the visual acuity of one eye is compromised - by occlusion (patching of the fellow eye); in this study, there is clear scope for confounding, as the amount of occlusion dose received is potentially highly influenced by child-specific factors. We use semiparametric additive linear mixed (SPALM) models as tools for estimating potentially non-linear covariate effects, and a balancing score approach to account for the effect of confounding/non-compliance.

1.1 *Quantifying dose-response over time: an analysis strategy*

Two related models are of interest when analyzing longitudinal data. The first (*absolute-level*) model assumes a repeated measures structure; each participant provides repeated time-varying covariate data over a number of intermediate measurements as well as a final, end-of-study measurement. The second (*interval-level*) model focuses on an interval-by-interval analysis, and takes as the response the change in response between successive measurements.

In this paper, we focus on the interval-level data, although the methods described could be employed in either context. With data in this form, a linear mixed-effects regression model provides a good means of initial analysis, with model selection (selection of influential covariates and interactions,

residual error correlation structures, random effects components) carried out using the Bayes Information Criterion (BIC). Potentially confounding relationships between covariates and treatment are next modelled flexibly using semiparametric additive linear mixed (SPALM) models (see, for example, Ruppert et al., 2003).

The regression strategy above takes no account of subject-controlled treatment level which is a potential source of bias in the estimation of treatment effect. Therefore, finally, we adopt the Generalized Propensity Score (GPS) (Hirano and Imbens, 2004), a balancing score approach for continuous treatment, to construct an instrumental variable that controls for this source of confounding.

The analysis methods are demonstrated in the context of the MOTAS data. First, a frequentist approach is taken. In the SPALM analysis, estimation of the covariate effects is carried out using maximum likelihood/REML methods. In the balancing score analysis, an approach utilizing plug-in versions of the balancing score to model the expected response for different dose levels is used. Secondly, the Bayesian versions of these methods are described and implemented; in the SPALM analysis, this necessitates the use of a novel prior structure on the mixed effects components, and in the balancing score analysis this facilitates a full assessment of uncertainty in the estimation under the specified model.

1.2 *Structure of Paper*

The paper is structured as follows: Section 2 provides details of our motivating example. Section 3 describes linear and semiparametric mixed models, and section 4, the balancing score approach. Bayesian versions of the

approaches are described in 5. The methods are discussed in section 6.

2. The Monitored Occlusion Treatment of Amblyopia Study (MOTAS)

2.1 *Amblyopia and its Treatment*

Amblyopia is the most common childhood vision disorder, affecting 1-5% of children. The condition is characterized by reduced visual functions, and usually affects only one eye. It has been associated with up to a three-fold increased lifetime risk of serious vision loss of the fellow eye (Rahi et al., 2002). A standard treatment for the condition is occlusion therapy, that is, patching of the functioning fellow eye. Perhaps surprisingly, 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, Moseley, Stephens and Fielder, 2004) was the first clinical study aimed at determining the dose-response relationship of occlusion. The study, for the first time, measured occlusion precisely using a dose monitor. We present the first comprehensive analysis of these data.

2.2 *Study Design and Implementation*

The MOTAS design and a full description of the study base have been published previously (Stewart, Fielder, Stephens and Moseley, 2002, Stewart et al., 2004). Prior to study entry, all children had a full ophthalmic assessment. Those who required spectacles entered the refractive adaptation phase; the remainder entered the occlusion phase directly. Children in refraction were prescribed to full-time spectacle use for 18 weeks, and scheduled for vision re-assessment at six week intervals. Children still considered

amblyopic began occlusion and were prescribed six hours of occlusion daily. That is, *the same dose* was assigned to all individuals in the occlusion phase of the study.

Visual acuity was measured on the logarithm of Minimum Angle of Resolution (logMAR) scale; improvement is indicated by a decrease in logMAR. Occlusion doses were recorded to the nearest minute by an *occlusion dose monitor* (ODM) (Fielder, Auld, Irwin, Cocker, Jones and Moseley, 1994). Visual function and monitored occlusion dose were recorded at approximately two-week intervals until acuity ceased to improve.

2.3 *Study Participants*

The study enrolled 116 children aged between 36 and 100 months, 29 of whom were excluded from the analysis as they did not meet the inclusion criteria or were lost to follow-up after a small number of clinic visits. Of the 87 patients that were suitable for inclusion in the statistical analysis, 15 saw their amblyopia resolved in the refractive adaption phase and did not enter the occlusion phase of the study. The remaining 72 were prescribed occlusion for six hours a day, but received different occlusion doses over different follow-up periods because of non-compliance. We assume that the loss to follow-up is not informative.

Profile plots of individual visual acuity trajectories over successive visits to the clinician are depicted in Figure 1. These indicate that a piecewise linear model of response is a reasonable foundation for our statistical models.

[Figure 1 about here.]

Time spent in occlusion was considered as a potential effect modifier, separate from the therapeutic effect of refraction or occlusion treatment,

or the developing age of the child whilst in the study. There also may be some adaption to the visual acuity testing procedures that could confound the relationship between treatment and visual acuity, and a time in study variable will be included in the model.

3. Linear and Semiparametric Mixed Model Analysis

Let the $N = 87$ patients in the study be indexed by i and the n_i+1 clinic visits by j, so that V_{ij}^a is the visual acuity for patient i on visit j at day t_{ij} into the study. Similarly, let D_{ij} be the (random) occlusion dose (in hours) observed in interval j; $D_{ij} = 0$ for the baseline observation in the occlusion phase. Let A_{ij} be the child's age in months at the start of interval j. Let L_{ij}, P_i and S_i denote the visual acuity at the start of interval, start of phase and start of study, respectively, and T_i denote the amblyopia type (anisometropic, mixed, strabismic), for patient i. For interval-level data, response is the change visual acuity in interval j between visit $j-1$ and visit j for patient i: $Y_{ij} = V_{ij}^a - V_{ij-1}^a$ for $j = 0, 1, ..., n_i,\ i = 1, 2, ..., N$.

3.1 *Linear Mixed Effects Models Analysis*

We begin with the assumption that variation in visual acuity has both a systematic and random component. Specifically, we assume an individual-specific random intercept at interval zero. That is

$$Y_{ij} = X_{ij}^{\mathsf{T}}\beta + \eta_i + \epsilon_{ij}, \tag{1}$$

and a presumed autocorrelation in the residual errors, ϵ, of a child.

3.1.1 Inference for the Linear Mixed Model Inference for the linear mixed model in equation (1) is achieved using penalized least-squares or likelihood

procedures under an assumption of Gaussianity of the residual errors; we give this version for ease of interpretation. Suppose, in conjunction with equation (3), $u \sim N(0, G)$ and $\epsilon \sim N(0, R)$ with u and ϵ independent. This model can be interpreted as $Y|\beta, u \sim N(X\beta + Zu, R)$, $u \sim N(0, G)$, yielding the marginal model $Y|\beta \sim N(X\beta, ZGZ^{\mathsf{T}} + R)$. Let $V = ZGZ^{\mathsf{T}} + R$. Then estimates can be found using the penalized maximum likelihood or the restricted maximum likelihood (REML), obtained by first integrating out β from the likelihood $Y \sim N(X\beta, V)$.

This model has a (model-based) Bayesian interpretation where the unknown parameters β and u are assigned independent prior distributions, with β having an improper uniform prior on $\mathbb{R}^{\text{ncol}(X)}$, and u assigned the Gaussian prior described above. In a fully Bayesian approach, G is set as a fixed hyperparameter, or assigned an informative prior distribution. Here, an empirical Bayes approach is used where G and the parameters in R are estimated using ML/REML. In section 5.1, the fully Bayesian approach is described, and the results of using a diffuse and an informative prior specification for G are compared.

3.1.2 Linear Mixed Model Results for MOTAS The model was fit in `R` using the `nlme` library. The BIC was used to select an optimal model. The optimal model for the refraction phase was $L + P + T$, and for the occlusion phase was $D + A + L + P + D.A + D.L + A.L$; residual plots raised no concerns about the fit of the model for the mean. Time on study, t, time in refraction, t_R, and time in occlusion, t_O, added little to the fit of the model.

Using the models and covariates identified above, residuals were explored

to help determine a suitable covariance structure amongst the repeated measurements. All models for the refraction and occlusion phases assume random intercepts at the individual level, as in equation (1), and for the occlusion phase, the addition of a random slopes model was also considered (see Diggle et al., 2001 for terminology). Autoregressive correlation structures were considered for the observed data, as in section 3.1; here, we retain the AR(1) model, where $\text{Corr}[Y_{ij}, Y_{ij'}] = \rho^{|j'-j|}$.

Using the BIC, it was apparent that a model with random intercepts and AR correlation provided a better fit to the data than the model without random effects; random slopes were not necessary. Table 1 gives parameter estimates for the terms in the final models using REML. The fit of this model yielded estimates of the residual error standard deviation and the correlation of $\hat{\sigma} = 0.0735$ and $\hat{\rho} = -0.1708$.

[Table 1 about here.]

In the refraction phase, visual acuity measurements suggests that prior to occlusion, the vision of anisometropic children given spectacles decreases on the logMAR scale (i.e. improves) on average by 0.085 (0.023,0.147) between each visit. Strabismic children exhibit a lesser degree of improvement while children of mixed type do not exhibit significant improvement. Children who were younger, and/or had higher logMAR at the start of occlusion and at the start of an interval all improved further for the same occlusion dose.

3.2 *Semiparametric Additive Linear Mixed Models Analysis*

We now fit a semiparametric additive linear mixed (SPALM) model to attempt to capture non-linearity in the covariate effects. The model we fit is

of the form

$$Y_{ij} = X_{ij}^{\mathsf{T}}\beta + \sum_{k=1}^{K} f_k\left(X_{ij}\right) + \epsilon_{ij}, \tag{2}$$

where the f_k, $k = 1, ..., K$, are functions of the covariates modelled semiparametrically; see, for example, Ruppert et al. (2003). We use the linear mixed model formulation,

$$Y = X\beta + Zu + \epsilon \tag{3}$$

where

$$E\begin{bmatrix} u \\ \epsilon \end{bmatrix} = \mathbf{0} \qquad Var[\theta] = \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix}$$

and the matrix X contains the fixed effects predictors, Z is the (basis function) design matrix in the semiparametric representation of the function of $f_1, \ldots, f_K$. Details of estimation procedures for this model are given in Web Appendix A.

3.2.1 Specification of the Semiparametric Design: The Truncated Spline Basis In the semiparametric additive model, the matrix Z contains the truncated spline basis terms, with columns corresponding to knots $\kappa_{k1}, \ldots, \kappa_{kM}$ for $k = 1, \ldots, K$. Typically, the random effects coefficients for function k are assigned a common Gaussian distribution so that the matrix G is diagonal; however this is not necessary (see Web Appendix B).

We use truncated spline basis functions to construct the semiparametric specification. Generically, for scalar x varying across a data-dependent range, we specify fixed (but data-dependent) knot positions $\kappa_{k1}, ..., \kappa_{kM}$, and model function f_k as

$$f_k(x) = \sum_{m=1}^{M} u_{km}(x - \kappa_{km})_+^q \tag{4}$$

where $u_{k1}, ..., u_{kM}$ are (random effects) coefficients for function k, and the basis function component $(x-\kappa_{km})_+^q = \max\{0, (x-\kappa_{km})^q\}$, so that a typical row of Z (an $N \times KM$ matrix) in equation (3) takes the form

$$[(x-\kappa_{11})_+^q \quad (x-\kappa_{1M})_+^q \quad \ldots \quad (x-\kappa_{KM})_+^q] \,.$$

We take $q = 1$ and use 10 knots at the covariate quantiles, with a knot also placed at zero, giving $M = 11$. For convenience, we transform (by translation) the covariates such that they are non-negative. The function f_k in equation (4) has vector of coefficients $\mathbf{u}_k$ of length M, which are assumed to be independent random effects with common variance σ_k^2, $k = 1 \ldots, K$. We also assume independence between $\mathbf{u}_1, \ldots, \mathbf{u}_K$, and thus retain a block diagonal structure for the entire random effect matrix.

The semiparametric model can be fit using `lme` in `R` for some choices of the residual error covariance R, and more generally using numerical procedures for general covariance specifications.

3.2.2 Semiparametric Mixed Model Results for MOTAS Three semiparametric components were used when considering MOTAS; a component for dose, D, a component for the interaction between D and (translated) Age at Interval, $A - 36$, and a component for the interaction between D and visual acuity at start of interval, $L + 0.175$. We used 10 fixed knots, with positions determined by covariate quantiles, but the results were robust to specifications with up to 50 knots. We examined three covariance structures for the repeated measures - an uncorrelated error model, an AR(1) model in interval number and an exponential decay-in-time covariance. We found that there was effectively no difference in the results in the resulting estimates of

the semiparametric components, nor in the BIC values for the two covariance models. Plots of the curves for the three predictors estimated under the uncorrelated error model are given in Figure 2.

[Figure 2 about here.]

3.3 *Comments on the Linear and SPALM analyses*

When looking for a dose-response effect, the associations between outcome and individual characteristics or treatment and these characteristics are not of interest in themselves. The splines used in SPALM models offer a much higher degree of adaptability to the data than linear models, and perhaps provide better control of confounding effects. The pattern of results from both the linear and the semi-parametric analyses of MOTAS are in line with ophthalmological opinion and practice: higher doses give greater improvement in vision and the impact is greater in children with inferior vision, but the effect is moderated by increasing age of the child.

A potential problem with the analyses of section 3 is that the (observational) study design makes interpretation of the results more complicated than it would be for an equivalent experimental study. In the next section we develop methods that address issues of confounding/non-compliance that are based on the Generalized Propensity Score (GPS) (Hirano and Imbens, 2004). While the GPS does not provide a causal parameter, it does allow us to construct an instrumental variable that facilitates meaningful inference on the treatment effect. We assert that if common results are found by applying two highly flexible but very different approaches, the analyst has likely eliminated bias due to non-compliance by effectively capturing confounding effects.

4. A Balancing Score Approach to Estimating the Dose-Response Relationship

The modelling approaches of section 3 do not recognize the observational nature of the occlusion dose received, raising the possibility that the effect of occlusion is misrepresented by the estimates presented due to confounding. It is plausible *a priori* that factors that influence the improvement in visual acuity in a treatment interval (for example, the age of the child, or current visual acuity) also affect the amount of occlusion dose received during the interval, which represents a classic confounded dependence structure.

To ascertain the true effect of dose, a causal analysis which accounts for the potential confounding between dose levels and other covariates is desirable. We present an approach approximating a causal analysis in this section. The principal tool used is the Generalized Propensity Score (see, for example Imai and Van Dyk, 2004, Hirano and Imbens, 2004) to account for possible confounding relationships between occlusion treatment and other covariates. We note here that the GPS does **not** provide a parameter that may be interpreted causally; see discussion below. The balancing property of the GPS may provide greater confidence that the analysis has successfully eliminated potential bias due to confounding.

4.1 *The Generalized Propensity Score*

We adopt the notation and terminology of causal modelling, and denote the response Y, occlusion dose D and covariates X. For child i in the study, we further denote $Y_i(d)$ for $d \in \mathcal{D}$ (the set of possible doses) as the set of *potential outcomes* that describe the dose-response function. Following Hirano and Imbens (2004) we define the (observed) Generalized Propensity

Score (GPS), $r(d, x)$ for dose d and covariate values x by

$$r(d, x) = f_{D|X}(d|x). \tag{5}$$

That is, $r(d, x)$ 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 defined by Rosenbaum and Rubin (1983) to continuous treatments.

Generically, the average causal effect of dose D is defined formally as the difference in expected outcomes for two dose levels d_0, d_1 for fixed covariate values $X = x$, that is

$$E[Y_i(D_i = d_1, X_i = x)] - E[Y_i(D_i = d_0, X_i = x)] \tag{6}$$

or the expectation of this quantity over the distribution of different X values in the study population. Our analysis does not return an estimate of this quantity, but instead returns an estimate of

$$E[Y_i(D_i = d_1, R_i = r)] - E[Y_i(D_i = d_0, R_i = r)] \tag{7}$$

where the GPS random quantity R_i acts as a balancing score, such that, given R, D and X are conditionally independent. Note that, in general, if $d_1 \neq d_0$, then we may have $R(d_1, x_1) = R(d_0, x_0)$ whether $x_1 = x_0$ or otherwise. However, equation (7) **does** facilitate consistent estimation of the dose-response relationship, as we may average each conditional expectations over the distribution of R if the balancing property holds. Whether the defined GPS has indeed the required balancing property (in that, within strata of approximately equal R values - based on, say, quintiles of the empirical

distribution of R - the conditional density value for $D = d$ does not depend on X) is a checkable assumption for any proposed scoring procedure.

The GPS quantity R acts as an instrumental variable in the regression for Y that, as equation (5) implies, is constructed by modelling D as a function of X. This yields a bias-removal strategy: we may 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. Practically, this implies a two-step strategy: build first a (regression) model to predict D given X that yields R, then a (regression) model for Y given D and R that has a built-in balancing of possible confounding between D and X that may bias the estimate of the effect of D on Y.

4.2 *Average and Conditional Potential Outcomes*

A key quantity of interest is the Average Potential Outcome (APO) at dose level d,

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

which traces the average dose-response relationship across covariate values for a given dose level. When considered as a function of d, this function can be interpreted as reflecting the causal relationship between dose and response.

The APO at $D = d$ can be estimated using the sample data; the estimation procedure depends on plug-in prediction of the dose effect. In section 5.2, we demonstrate how a simple Bayesian procedure can be used to propagate uncertainty through the model in a coherent fashion.

The approach to estimating the APO at dose level d, $\mu(d)$, is outlined by Hirano and Imbens (2004) and summarized in Web Appendix C. This

procedure produces an approximation to the expectation of the potential outcome at a specified dose value d; the "average" is over the distribution of covariate values. A similar procedure can be used to approximate the expected *conditional* potential outcome (CPO), $\xi(d, x)$, at a specified pair (d, x) a quantity that may be of equal interest. Furthermore, bootstrap procedures that re-sample with replacement from $(x_1, \ldots, x_n)$ can be used to produce uncertainty bounds for $\mu(d)$, and for $\xi(d, x)$ if sufficient replicates at predictor values $X = x$ are available.

4.3 *Applying the GPS to the MOTAS data*

4.3.1 The GPS model In the MOTAS study, dose is a continuous variable, so the GPS approach is appropriate. However, even in the occlusion phase, a non-negligible proportion of intervals (63 out of 411, around 15%) had a corresponding zero dose. The predictive model $f_{D|X}(d|x, \beta)$ must acknowledge the mixture nature of the dose distribution, so we assume that, given $X = x$,

$$D \stackrel{\mathcal{L}}{=} \pi(x, \gamma)\delta_{d=0} + (1 - \pi(x, \gamma))D_+$$

where $\delta_{d=0}$ is a unit mass on zero, D_+ is a strictly positive random variable whose distribution depends on $X = x$ and β, and $\pi(x, \gamma)$ is a mixing weight depending on $X = x$ and parameters γ. Estimation in this model is straightforward when a parametric distribution is used for D_+; we select the Weibull model, with scale parameter modified by X, although the log-logistic or log Gaussian distributions are suitable two parameter alternatives. To estimate γ, we fit a logistic regression model to the binary $(D = 0/D > 0)$ dose data. Any such general regression model that induces a balancing property can be used; the Weibull-based model is convenient, and, it transpires, suitable.

On exploring possible dose/covariate relationships suitable for inclusion in the GPS model, we discovered that time in occlusion was influential in predicting $D = 0$. In the logistic regression model, $\text{logit}\{P[D = 0]\} = \gamma_0 + \gamma_1 t_O$, the estimates (standard errors) for the two parameters were $\widehat{\gamma}_0 = -2.632(0.244)$ and $\widehat{\gamma}_1 = 7.282\text{e-}3$ (1.398e-3) respectively. In the Weibull model, interval number, visual acuity at start of interval, L, length of interval (in days), and amblyopic type, T, were all useful in explaining the variability in D_+. The first three of these variables had coefficient estimates (s.e.) in a Weibull proportional hazards model of 0.036(0.014), 0.537(0.175) and 6.68e-3 (3.21e-3) respectively. Thus, we have for the (estimated) GPS

$$\widehat{r}(d,x) = \begin{cases} \widehat{\pi}(x,\widehat{\gamma}) & d = 0 \\ (1-\widehat{\pi}(x,\widehat{\gamma}))f(d|x,\widehat{\phi},\widehat{\beta}) & d > 0 \end{cases}$$

where

$$f(d|x,\phi,\beta) = \frac{\phi d^{\phi-1}}{\exp\{\phi x^{\mathsf{T}}\beta\}} \exp\left\{-\frac{d^{\phi}}{\exp\{\phi x^{\mathsf{T}}\beta\}}\right\} \quad d > 0.$$

For the GPS to act as a balancing score, the distribution of D should not depend on X within strata of R. A check of the distribution within quintile categories of $\widehat{r}$ reveals no apparent relationship between D and X.

4.3.2 The Observable Model The predictive model for change in visual acuity, Y, is required as a function of occlusion dose, D, and the GPS, R. In Hirano and Imbens (2004), only the expected value of Y given D and R is considered; we follow that strategy here, modelling

$$E_{Y|D,R}[Y|D = d, R = r(d,X), \alpha] = \alpha_0 + \alpha_1 d + \alpha_2 r + \alpha_{12} d.r \qquad (8)$$

and estimate the parameters using least-squares, although a model assuming independent Gaussian residual errors appears sensible and facilitates a

likelihood analysis. The model for the expectation in equation (8) could be replaced by a piecewise constant "partition" model defined on a tesselation of the (D, R) space without complicating the inference to any great degree.

Using the model in equation (8), we obtain $\widehat{\alpha}_0$ = -1.865e-2 (1.195e-2), $\widehat{\alpha}_1$ = -2.744e-4 (1.078e-4), $\widehat{\alpha}_2$ = 3.478e-2 (6.323e-2) and $\widehat{\alpha}_{12}$ = -9.247e-2 (4.104e-2) respectively. Thus, the empirical average over the 411 interval observations, using estimated (potential) GPS values, $\widehat{r}(d, X_i)$, returns a consistent estimate of the average potential dose-response for each d.

4.3.3 Results A plot of the dose-response curve is presented in Figure 3. This plot indicates that the association between dose and visual acuity, when confounding between dose and the covariates is adjusted for using the GPS approach, is appreciable; the average potential effect on logMAR measurement Y is significantly negative (corresponding to vision improvement) over the entire range of positive doses considered. A numerical summary for the (frequentist) APO is found in Table 2. The GPS analysis gives an estimated effect of dose that is somewhat less than that found by the SPALM analysis, although direct comparisons are difficult.

[Figure 3 about here.]

The model used, for example, in equation (8) can be readily extended to a more flexible model; here, qualitative changes in the inferences made after the inclusion of quadratic terms in the linear predictor were minimal.

[Table 2 about here.]

5. Bayesian Approaches

The results of the likelihood-based analysis above have identified key predictors in the model for changes in visual acuity and quantified the influence of occlusion does on improvement in vision. However, the models are complex and often require plug-in procedures. Therefore, we implemented a Bayesian analysis that propagates the uncertainty in the inference in a fully coherent fashion. We perform inference using Markov chain Monte Carlo. In the interests of brevity, Bayesian linear mixed model results are omitted in order to devote greater attention to the novel aspects of the analysis: the Bayesian SPALM model and the Bayesian Generalized Propensity Score. Furthermore we omit details of the posterior summaries and focus only on the semiparametric components in the SPALM model.

5.1 *Bayesian Semiparametric Modelling Analysis and Results*

A semiparametric model similar to the one described in section 3.2 can be fitted in the Bayesian framework. Most importantly, the model is fundamentally unchanged from that described in section 3.2; the principal difference in the model specification is the prior used for the random effects coefficients that are used to construct matrix G. In an initial analysis, a non-informative prior specification is used, where the diagonal elements of G are set to be 1.0e+10. The results for this prior specification, the empirical Bayes specification implied by the classical analysis, and the analysis based on an informative prior specification (Web Appendices B and D) are depicted in Figure 4. Qualitatively, the differences between the prior specifications, and the interaction between dose and the other predictors are evident. It is clear, for example, that the empirical Bayes prior appears to shrink the magnitude of

effect more towards zero relative to the informative prior. Figure 5 depicts the Bayesian estimates and 95% credible interval for the informative prior.

[Figure 4 about here.]

[Figure 5 about here.]

5.2 *The Bayesian Generalized Propensity Score*

The analysis of section 4 can be recast in a Bayesian framework that retains the dual regression model aspect, but uses a fully Bayesian procedure to report the uncertainty in the estimated APO function via its posterior distribution.

Recall the components of the GPS model, namely the two conditional densities that give the model specification, $f_{D|X}(d|x, \gamma, \beta)$ and $f_{Y|D,R}(y|d, r, \nu)$. We obtained ML estimates for γ, β and ν, and then used a plug-in approach to estimate the APO function $\mu(d)$ for a range of values of d. The parallel Bayesian procedure replaces the plug-in approach with an approach that averages over the posterior distribution of the unknown parameters. For typical model specifications, MCMC-based procedures are straightforward; due to the model structure, conditional updating of the model parameters in the two stages of the model given the observed data is readily achieved (see Web Appendix E).

5.2.1 Obtaining the Bayesian APO and CPO A suitable quantity to report in the Bayesian framework is the posterior predictive expected response, pointwise for different values of d. For the Bayesian CPO, this computation is done for a fixed value of x, whereas for the APO, we must average the responses over the distribution of all possible values of X. In the

model described in section 4, the conditional expected value for Y, given $(D, R) = (d, r)$ and α is available directly as the Gaussian mean. Computing r given d, x and (β, γ) is straightforward. It also is straightforward to produce a sampled value of the conditional expectation at iteration m, given any d, x and $(\alpha^{(m)}, \beta^{(m)}, \gamma^{(m)})$, and appeal to Rao-Blackwellization to obtain the expectation with respect to the joint posterior distribution. When such analytic results are not available, computation of the expected response can be carried out in a number of ways; the simplest method involves Monte Carlo approximation. For the Bayesian APO, computing the expectation over the distribution of covariate values is more problematic but achievable using bootstrap methods; for MCMC iteration m, we carry out bootstrap re-sampling of the covariate values, obtaining x^{res_b} for $b = 1, \ldots, B$, and compute the conditional expectation

$$\widehat{\mu}^{(m)}(d) = \frac{1}{B} \sum_{b=1}^{B} E[Y|d, x^{res_b}, \alpha^{(m)}, \beta^{(m)}, \gamma^{(m)}].$$

This quantity can then be averaged over the MCMC iterations to produce the Bayesian APO. For the CPO, conditioning on a single given x value is more straightforward. The Bayesian approach yields the entire posterior distribution for the CPO, which can be summarized via quantile summaries for, say, 95% credible intervals at different values of d.

5.2.2 Results An MCMC-based Bayesian analysis of the Weibull-mixture/Gaussian model described in section 4.3 was implemented in `R`. We again focus only on the APO, using the form in equation (8) as the model for mean effect. The results are depicted in Figure 3, with the frequentist APO from section

4.3 included for comparison. A numerical summary is provided in Table 2. In this example, there appears to be little qualitative difference between the frequentist and Bayesian inferences.

6. 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. Semiparametric additive linear mixed (SPALM) models and Generalized Propensity Scores are of practical value: each is tractable and flexible, particularly when implemented in a Bayesian framework. We presented Bayesian versions of the SPALM model and of the GPS, and implemented the GPS with a mixture model.

In this paper, we quantified the dose-response relationship between improvement in vision and occlusion of amblyopic children. Our analysis is just one example where the relationships between dosing and covariates are complex and not well-approximated with linear forms. We have presented a cohesive framework to examine the impact of a continuous treatment given longitudinally with patient non-compliance.

An area of ongoing statistical research and for future ophthalmic research is that of optimal dynamic regimes. Dynamic regimes allow treatment to change over time based on patient history up to the current time - for example, an optometrist may prescribe a tapering of occlusion as visual acuity improves over time. A number of innovations in the design and analysis of dynamic regime trials have recently been made (see, for example, Murphy, 2003; Robins, 2004; Lavori and Dawson, 2004; Moodie, Richardson and

Stephens, 2006); an implementation in relation to MOTAS can be obtained through author Moodie.

7. Supplementary Materials

The Web Appendices referenced in sections 3.2, 4.2, 5.1, and 5.2 are available under the Paper Information link at the Biometrics website http://www.tibs.org/biometrics.

Acknowledgements

Moodie was supported by a Merck Company Foundation BARDS Fellowship. MOTAS was supported by The Guide Dogs for the Blind Association, UK.

References

Diggle, P. J., Heagerty, P., Liang, K.-Y. and Zeger, S. L. (2001). *Analysis of Longitudinal Data*. Oxford University Press, 2 edition.

Fielder, A. R., Auld, R., Irwin, M., Cocker, K. D., Jones, H. S. and Moseley, M. J. (1994). Compliance monitoring in amblyopia therapy. *Lancet* **343**, 547.

Hirano, K. and Imbens, G. W. (2004). The propensity score with continuous treatments. In Gelman, A. and Meng, X.-L., editors, *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, New York. John Wiley.

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

Lavori, P. W. and Dawson, R. (2004). Dynamic treatment regimes: practical design considerations. *Clinical Trials* **1**, 9–20.

Moodie, E. E. M., Richardson, T. S. and Stephens, D. A. (2006). Demystifying optimal dynamic treatment regimes. *Biometrics* **360**, 597–602.

Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society, Series B* **65**, 331–366.

Rahi, J. S., Logan, S., Timms, C., Russell-Eggitt, I. and Taylor, D. (2002). Risk, causes and outcomes of visual impairment after loss of vision in the non-amblyopic eye: a population-based study. *Lancet* **360**, 597–602.

Robins, J. R. (2004). Optimal structural nested models for optimal sequential decisions. In Lin, D. Y. and Heagerty, P., editors, *Proceedings of the Second Seattle Symposium on Biostatistics*, New York. Springer.

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

Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press.

Stewart, C. E., Fielder, A. R., Stephens, D. A. and Moseley, M. J. (2002). Design of the monitored occlusion treatment of amblyopia study (MOTAS). *British Journal of Ophthalmology* **86**, 915–919.

Stewart, C. E., Moseley, M. J., Stephens, D. A. and Fielder, A. R. (2004). Treatment dose-response in amblyopia therapy: the monitored occlusion treatment of amblyopia study (MOTAS). *Investigations in Ophthalmology and Visual Science* **45**, 3048–3054.

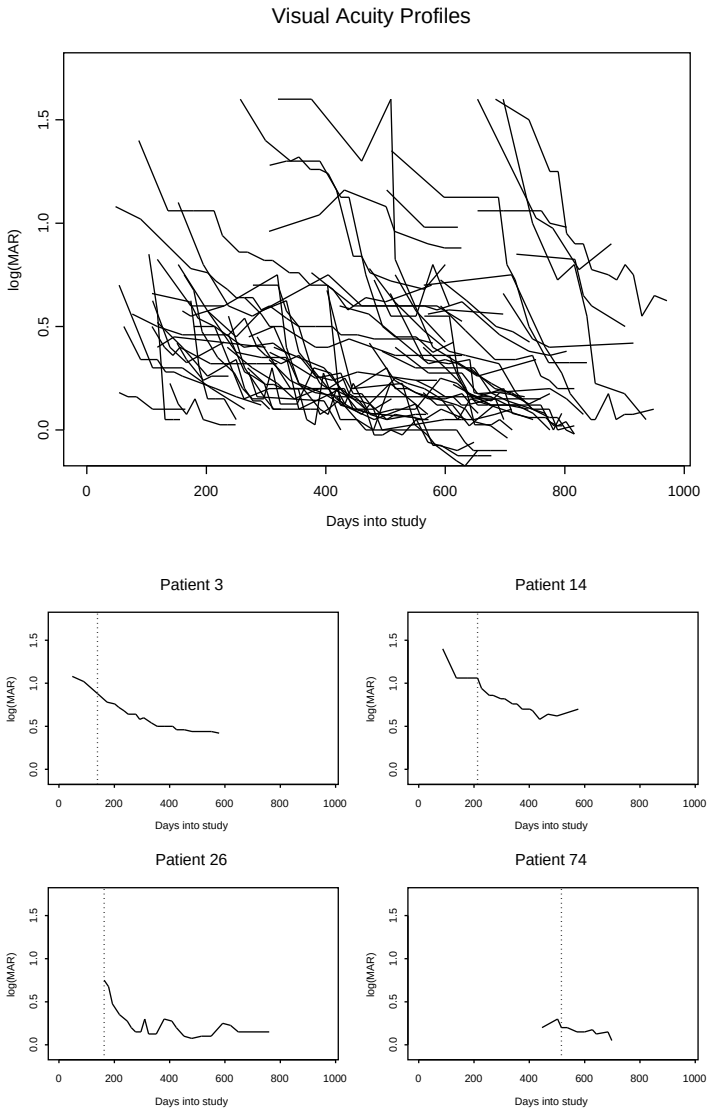

Figure 1. Profile plots for the individuals in the MOTAS study (top) and for four selected patients, with the start of occlusion indicated by the dotted line (bottom).

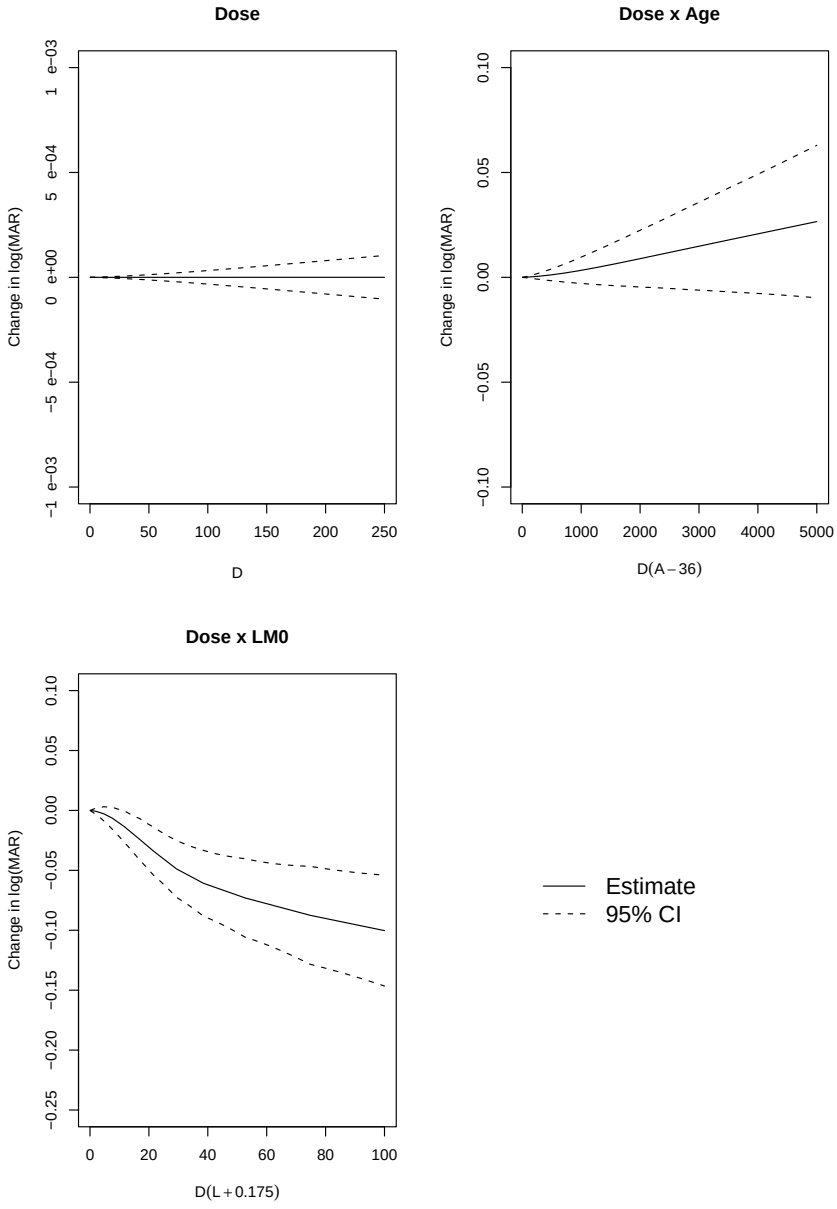

Figure 2. Estimated semiparametric functions from the maximum-likelihood SPALM analysis for uncorrelated error model, with pointwise 95% confidence interval.

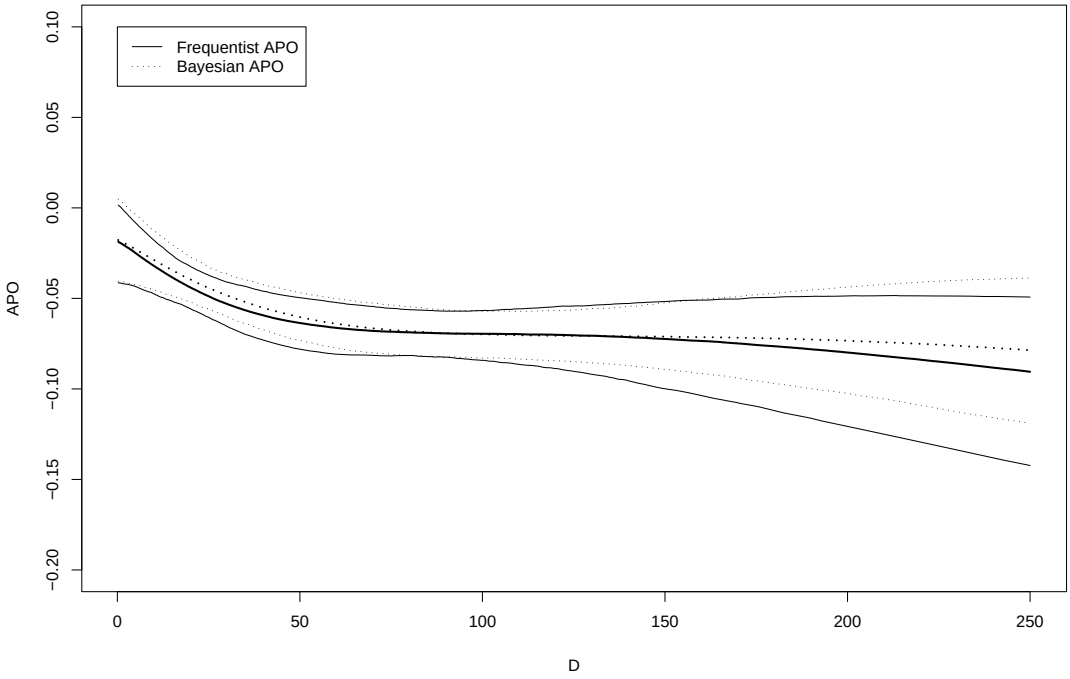

Figure 3. The estimated average potential effect for doses in the range 1 to 250 hours per interval (with approximate pointwise 95% credible interval) in the parametric Bayesian causal analysis from section 5.2. Estimated frequentist APO and 95% interval included for comparison

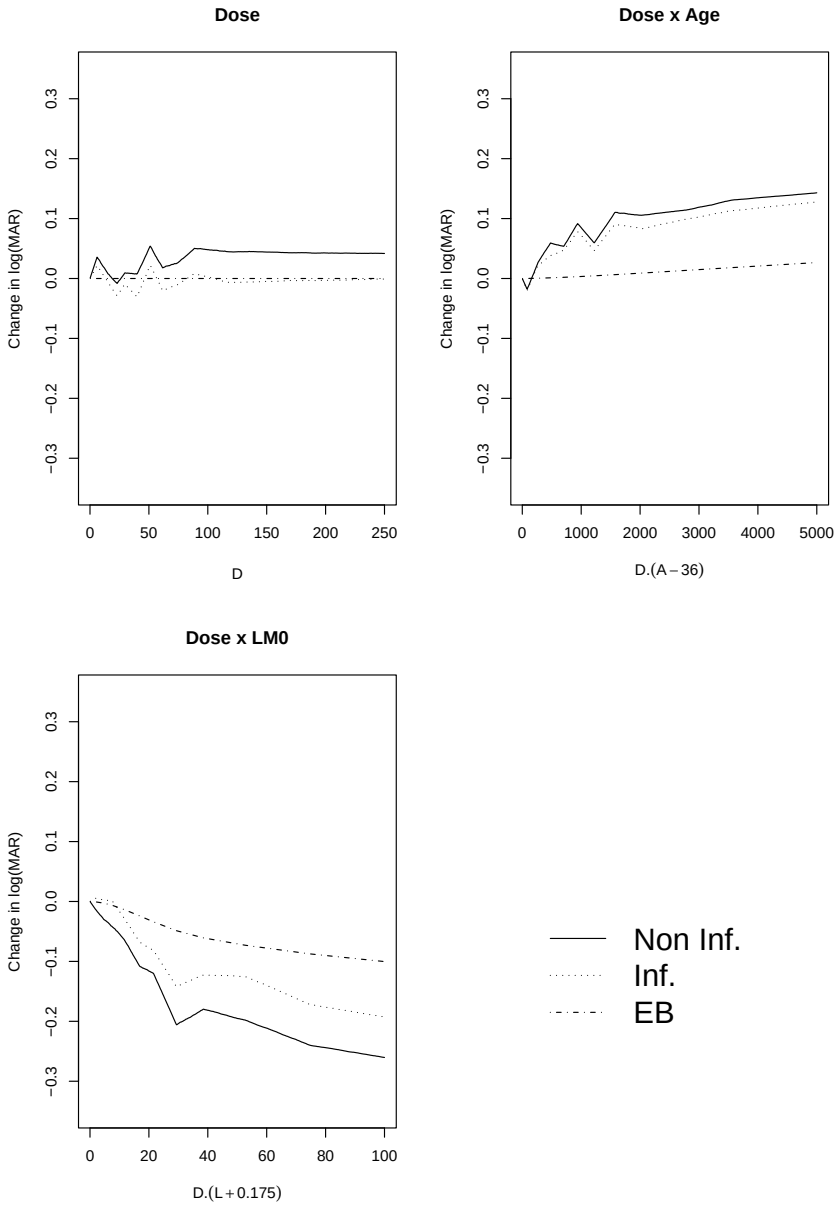

Figure 4. Bayesian median estimates under the three prior structures: non-informative, informative and empirical Bayes

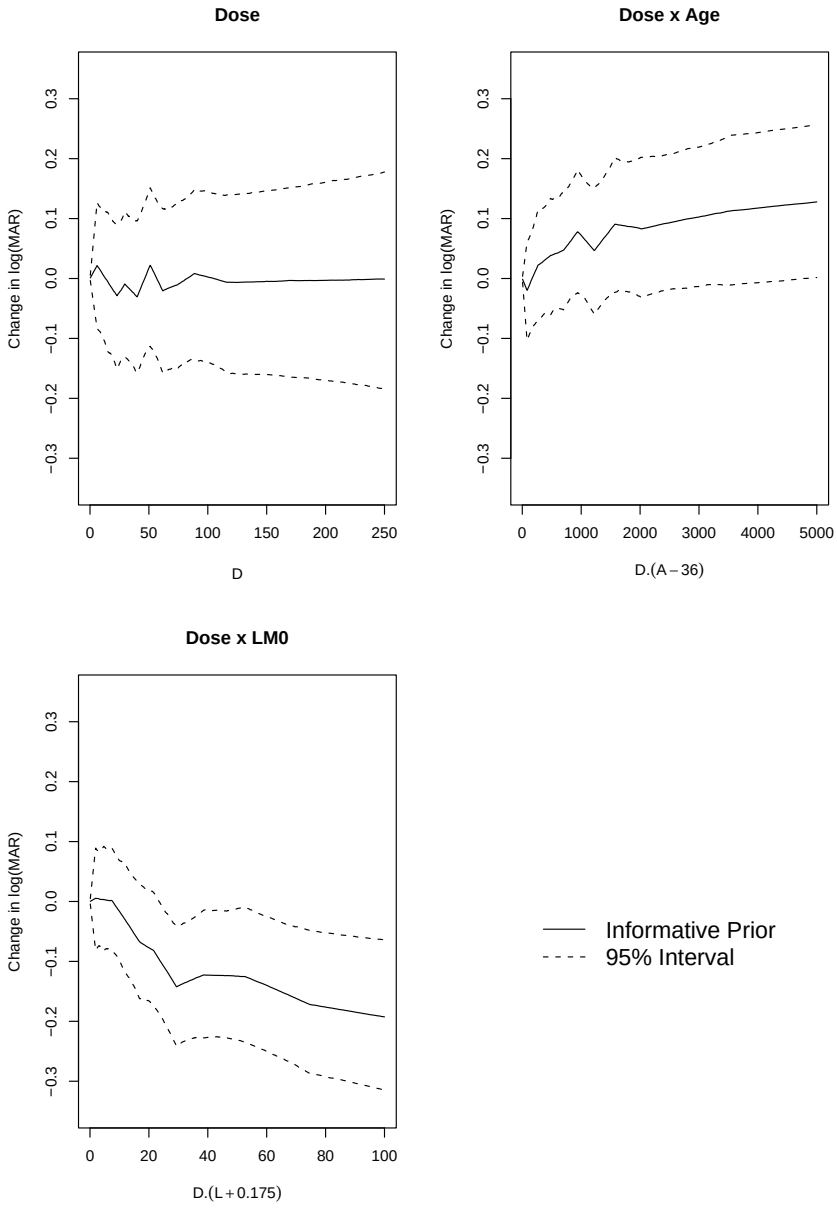

Figure 5. Bayesian median estimates and 95% credible interval for the informative prior.

Table 1

Estimates and standard errors for the parameters from a random intercepts linear mixed effects model with AR correlation.

Phase	Term	Est.	s.e.	t.stat	p
Refraction	Int.	-8.509e-2	3.149e-2	-2.702	0.007
	L	-6.376e-1	4.145e-2	-15.383	0.000
	P	4.839e-1	5.405e-2	8.952	0.000
	T_M	8.313e-2	3.640e-2	2.284	0.023
	T_S	1.682e-2	3.818e-2	0.441	0.660
Occlusion	Int.	-5.728e-3	3.473e-2	-0.165	0.869
	D	-8.645e-4	3.226e-4	-2.680	0.008
	A	-5.738e-4	4.920e-4	-1.166	0.244
	L	-5.113e-1	5.557e-2	-9.237	0.000
	P	1.234e-1	2.144e-2	5.755	0.000
	$D.A$	1.339e-5	5.063e-6	2.645	0.008
	$D.L$	-9.705e-4	2.959e-4	-3.279	0.001
	$A.L$	4.912e-3	9.242e-4	5.315	0.000

Table 2

Summaries of the APO (on the logMAR scale) for changing dose amount per interval: 5000 bootstrap or posterior samples.

		Dose (hours)					
	Quantile	25	50	75	100	125	150
Frequentist	0.025	-0.0604	-0.0780	-0.0818	-0.0842	-0.0903	-0.0999
	0.250	-0.0528	-0.0683	-0.0729	-0.0743	-0.0766	-0.0804
	0.500	-0.0488	-0.0636	-0.0684	-0.0695	-0.0704	-0.0724
	0.750	-0.0449	-0.0589	-0.0638	-0.0651	-0.0649	-0.0651
	0.975	-0.0371	-0.0496	-0.0555	-0.0567	-0.0542	-0.0517
Bayesian	0.025	-0.0559	-0.0731	-0.0808	-0.0829	-0.0850	-0.0892
	0.250	-0.0480	-0.0643	-0.0717	-0.0741	-0.0755	-0.0772
	0.500	-0.0443	-0.0603	-0.0675	-0.0699	-0.0706	-0.0711
	0.750	-0.0403	-0.0561	-0.0633	-0.0658	-0.0658	-0.0649
	0.975	-0.0326	-0.0469	-0.0537	-0.0570	-0.0564	-0.0524