Quantifying the Efficacy of Occlusion Therapy in the Treatment of Amblyopia

Erica E. M. Moodie
Department of Biostatistics, University of Washington, Seattle.
eemm@u.washington.edu

David A. Stephens
Department of Mathematics, Imperial College London
d.stephens@imperial.ac.uk

Catherine E. Stewart, Alistair R. Fielder, Merrick J. Moseley
Department of Optometry and Visual Science, City University, London

Abstract

Amblyopia is the most common visual disorder of childhood, yet the contributions of the two principal treatments (spectacle wear and occlusion) to outcome are unknown. The MOTAS study investigated the dose-response relationship between occlusion (patching) and improvement in visual acuity. The use of the Occlusion Dose Monitor (ODM) to record the amount of occlusion dose received by the child represents a major innovation and is unique to this study. Key issues in the analysis are to identify the pattern of covariate influence and to quantify the degree to which occlusion dose proves therapeutic in treating amblyopia. We present mixed model, semiparametric and Bayesian analyses, and discuss the potential impact of different dosing strategies in future studies. We also explore causal inference approaches, based on a Generalized Propensity Score (GPS) balancing strategy, in an attempt to estimate the true (causal) effect of dose on improvement in visual acuity and introduce a Bayesian version of the GPS that is readily implementable in a simulation-based analysis.

KEYWORDS: Amblyopia, Occlusion Therapy, Semiparametric modelling, Causal Inference, Bayesian Analysis.

1 Introduction

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 involves patching of the functioning fellow eye. However, perhaps surprisingly, the apparent beneficial effect of occlusion therapy has never been accurately quantified, partly because the accurate measurement of the occlusion dose received by a child has proved difficult. This paper describes a statistical analysis of data from a recent study of amblyopic children, with the goals of quantifying the magnitude of the treatment effect and characterizing the dose-response relationship. This study was ground-breaking at its inception, as, for the first time, the occlusion dose that each subject received during the study was to be measured precisely using a dose monitor. We present the first comprehensive analysis of the study data. An issue of particular interest is to uncover the true effect of occlusion, when other factors may also lead to perceived improvement in vision.

1.1 Amblyopia: History of the condition and its treatment

Research in the 1960s and 1970s demonstrated that the developing visual system is highly sensitive to deprivation. This led to the concept of a visual sensitive period, ending at around 6 to 7 years, which results in amblyopia if disrupted. The clinical implication of this research was that amblyopia should be both identified and treated in early childhood. This critical notion has influenced health service management in many countries, so that in the UK at least, national screening for amblyopia is recommended between 4 and 5 years and overall (Hall (1989)). About 90% of children's eye services work is amblyopia-related.

The principal treatment for amblyopia for over 250 years has been *occlusion* (*patching*) of the better-seeing (*fellow*) eye by an opaque patch. Occlusion therapy has long been believed to be beneficial in young children, however optimal dosing strategies are still unknown. A recent clinical study (Repka et al. (2005)) that followed children under 7 for two years to compare patching to atropine eye drops saw comparable improvements (3.7 vs. 3.6 lines) in visual acuity between the groups. A study was also conducted to examine the impact of prescribing occlusion for six vs. all waking hours each day for children under 7 with moderate amblyopia (PEDIG (2003)). At 4 months, there was no significant difference in the improvements in visual acuity of children in each randomization group, however (reported) concordance to prescribed hours of occlusion was significantly lower in children prescribed all-day occlusion.

The Monitored Occlusion Treatment of Amblyopia Study (MOTAS) (Stewart et al. (2004)) was the first clinical study aimed at determining the dose-response relationship of occlusion therapy. Importantly, the effects of refraction were differentiated from those of occlusion and concordance (compliance) with treatment was measured objectively. The study is described in detail in section 2.

1.2 The Aim of this Paper

The aim of this paper is to establish the dose-response relationship between occlusion treatment and improvement in visual acuity, and to understand how this relationship is modified by patient characteristics. Previous analyses of these data have been expository and exploratory. For example, in Stewart et al. (2004), the total dose vs. total improvement relationship was computed using semiparametric regression and bootstrap methods over the whole patient cohort. In this paper, we study the dose-response relationship to gain a quantitative understanding of the potential therapeutic benefits of occlusion therapy.

In the next section, we give more details of the MOTAS study design and outline the strategy behind our analyses in section 2.3. In section 3 we describe a linear mixed model analysis accounting for the repeated measures nature of the data. An interval-by-interval analysis follows; the results are presented in section 3.3.1. The causal analysis using the Generalized Propensity Score is presented in section 4. Bayesian modelling and inference can be found in 5. A Bayesian analysis of dosing strategies is shown in section 5.6 and a discussion of the implications for optimal treatment strategies is given in section 6.

2 The MOTAS study

The MOTAS study comprised two phases: *refractive adaptation* (or *refraction*) and *occlusion* (see section 2.1). Profile plots, tracking individual visual acuity trajectories over successive visits to the clinician, are depicted in Figure 1. Acuity measurements are on the logarithm of Minimum Angle of Resolution (logMAR) scale; improvement is indicated by a decrease in logMAR. Figure 1 implies a general beneficial effect of occlusion therapy, as the majority of individuals have generally downward sloping trajectories. However, the amount of occlusion dose within each interval is not reflected in this plot. In the lower panel, profiles for four selected patients are depicted, with the beginning of occlusion phase indicated.

The MOTAS study was revolutionary as for the first time the amount of occlusion therapy that each child received was accurately recorded using an electronic monitor. This pioneering study offered the possibility of quantification of the true magnitude of the dose effect in the improvement in vision, accounting for the possible influence of other child-specific factors such as the age of the child.

2.1 Study Design and Implementation

The MOTAS study design and a full description of the study base have been published previously (Stewart et al. (2002, 2004)). Prior to study entry, all children had a full ophthalmic assessment. Those who required spectacle correction entered the refractive adaptation phase, and the remainder entered the occlusion phase directly. Children in refraction were prescribed to full-time spectacle-wear for eighteen weeks, and scheduled for vision re-assessment at six-weekly intervals. Children still considered amblyopic began occlusion and were prescribed six hours occlusion per day.

Occlusion doses were recorded to the nearest minute by an *occlusion dose monitor* (ODM) (Fielder et al. (1994)). The ODM consists of an eye patch with two small electrodes attached to its under-surface connected to a battery-powered data logger. At each visit, data from the ODM was downloaded to a PC and parents were given the opportunity to review their child's concordance. Both visual function and monitored occlusion dose were recorded at approximately two-weekly intervals until acuity ceased to improve; the decision to end occlusion was made in a pragmatic fashion, after two inflexions in an acuity against time plot or identical acuity measurements on three consecutive visits. Participants returned to standard clinical care on completion of the occlusion phase.

The study enrolled a total of 116 children aged between 36 and 100 months over a period of two years. Of these, 87 were suitable for inclusion in the statistical analysis. Fifteen did not enter the occlusion phase, but had their visual impairment successfully treated with refractive adaptation - these children are included in the analysis. The remaining 72, although prescribed occlusion for six hours a day, received different occlusion doses over different follow-up periods. The 29 children excluded from the analysis were either deemed not to meet the inclusion criteria or lost to follow-up after a small number of clinic visits. We will assume that the loss to follow-up is not informative, and that the 87 patients entering the analysis are a suitable random sample from the amblyopic population.

2.2 The Lack of Randomization and a Control Group

In the MOTAS study, it was not possible to implement a randomized controlled design for two important reasons. First, it was not considered ethically justified to deny any child the possibility of obtaining a maximum possible beneficial dose. Secondly, it was not possible to propose a sensible and implementable designed experiment, as the level of concordance for any child was not known before the experiment was started. Despite these ethical and practical difficulties, randomized studies of occlusion therapy in amblyopia have recently been approved under strictly regulated protocols (see, for example, PEDIG (2003)). Indeed, the results of the MOTAS study have facilitated a randomized trial (ROTAS), carried out by the MOTAS cooperative, which has recently been completed; see Stewart et al. (2004) for further details.

2.3 Analysis Strategy

The plots in Figure 1 indicate that a piecewise linear model of response may be a sound foundation for the statistical modelling on which we embarked. Two related models are of interest. The first model assumes a repeated measures structure; each child in the study undergoes repeated assessment of their visual acuity over a number of clinic visits. We term the data in this form the *absolute-level* data. The second model focuses on an interval-by-interval analysis, and takes as the response as the change in visual acuity between successive clinic visits. Data in this form will be termed the *interval-level* data. In the study, the ODM data are available on a by-dose basis, that is, the ODM records each separate interval during which the child is wearing the monitor. In this paper, we restrict attention to simple functions of the total dose (in minutes) and the length (in days) of the interval between clinic visits. Elementary numerical summaries imply that an increasing dose within an interval yields an improvement in visual acuity; the mean (standard error) changes in visual acuity in an interval for doses in the range 0-50, 50-100, 100-150, 150-200 and above 200 minutes are -0.0327 (0.006), -0.0744 (0.009), -0.101 (0.020), -0.0430 (0.026) and -0.0512 (0.019) respectively. However, the nature of the dose-response relationship is unclear.

In our analysis, time spent in the study (or 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 may be a degree of adaption to the visual acuity testing procedures that is a potential confounder for treatment, and so including a suitable time in study variable in the model will detect the effect of these two outcome modifiers.

3 Linear Mixed Model and Semiparametric Analysis

Let the $N = 87$ patients in the study be indexed by i and the n_i+1 clinic visits by j, so that Y_{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; for those patients who enter the study in the refraction phase, but do not require occlusion, $D_{ij} = 0$ is identically zero for all j. Similarly, $D_{ij} = 0$ for the first observation in the occlusion for those children who entered that phase. Let A_{ij} be the child's age in months at the beginning 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 these data, it is preferable to use cumulative dose for that individual up to visit j, D_{ij}^c, as a time-varying covariate.

In this preliminary model-fitting stage, a number of models and model refinements that employ various combinations of the variables identified above were considered. The most complex model that we study contains the following main effect covariates: Amblyopia type, T, Visual Acuity at Start of Study, S, Age at start of interval, A, Cumulative Occlusion dose, D^c, Time in Refraction, t_R, and Time in Occlusion, $t_O = \max\{0, t - t_0\}$, where t_0 represents the start of occlusion, for those children that enter occlusion. We also fit interactions between D^c and A, D^c and S, D^c and t_O, t_O and A, and t_O and S. Higher order interactions were not found to improve the fit of the model.

3.1 Analysis of the Absolute-Level Data

3.1.1 A Multivariate Gaussian linear model

In a preliminary analysis, a standard Gaussian linear model was fit to the set of 684 observations, without any by-individual grouping was fit using all the covariates listed above, and a selection of second order interactions. This model had an adjusted R^2 value of 0.918; the estimated residual error variance is 0.021 ($\sigma = 0.144$). A residual plot revealed that, apart from a small number of outlying observations, the model fit was adequate. Using a Bayes Information Criterion (BIC), we constructed an optimal model nested inside the global model indicated above. The optimal model under BIC for the refraction phase was $t_R + T + S$, and for the occlusion phase was $t_O + S + A + D^c + D^c.A + D^c.t_O + t_O.S$; the overall BIC measure was -626.442, and a residual plot revealed that this model was also adequate. The R^2 value was 0.918, with $\hat{\sigma} = 0.145$. The estimated coefficients in the model confirm conventional ophthalmological wisdom; for cumulative dose, D^c, the interaction between dose and age at interval, $D^c.A$, and the interaction between dose and time in occlusion, $D^c.t_O$, the estimates (standard errors) are -9.975e-4 (1.756e-4), 1.121e-5 (2.648e-6) and 1.517e-6 (3.351e-7) respectively. Thus it appears that visual acuity improves with increasing cumulative dose, although this improvement is moderated by the age of the child and the time spent in occlusion.

A multivariate Gaussian structure for the longitudinal data is more immediately appropriate. For individual $i = 1, ..., N$, $V_i \sim N_{n_i+1}(X_i\beta, R_i)$, where X_i is a $(n_i + 1) \times p$ covariate matrix, β is a $p \times 1$ coefficient vector, and R_i captures the covariance in the random observation errors ϵ_i. The parametric form of covariance structure used here is the stationary exponential decay model, where $\mathrm{Cov}[V_{ik}, V_{il}] = \lambda \exp\{-|t_{ik} - t_{il}|^\nu / \zeta^\nu\}$ with common parameters $\lambda, \zeta, \nu > 0$ for all individuals in the study. This model uses time into study, t_{ij}, in the metric to define the degree of covariance.

Estimates and standard errors derived from the multivariate Gaussian model under maximum likelihood (ML) and restricted maximum likelihood (REML) criteria for the estimation of the observation covariance matrix are available numerically, and are presented in Table 1. The standard errors are slightly larger under REML, but there is no substantive difference in the inference for either model. The multivariate models that account for the repeated measures nature of the data are in general much preferred under BIC to models with independent errors (BIC is now -1057.044). A residual plot revealed that the fit of this model is satisfactory.

3.1.2 A Linear Mixed Effects Model

We now assume that the variation in visual acuity has both a systematic and random component. Specifically, we use a random intercepts model, where there is an individual-specific random

intercept at interval zero. That is

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

and a presumed autocorrelation in the residual errors ϵ. The model was fit in `R` using the `nlme` library. The presumed correlation structure specified in `R` for the grouped data, and chosen by BIC is an AR(1) structure in visit number; this structure ignores the calendar time between visits, but no improvement was found in the fit when a continuous time model was fit. When refit using ML, the optimal model had BIC measure -1086.919. Most importantly, the coefficients of covariate terms including cumulative dose were significantly different from zero in a Wald test; the estimates (s.e.) were -1.325e-3 (2.316e-4), 1.213e-5 (3.362e-6) and 1.742e-6 (3.617e-7) for D^c, $D^c.A$ and $D^c.t_O$ respectively. The correlation structures available in the `nlme` package in `R` are limited. By experimentation, we found that the `R` continuous AR(1) model, with autocorrelation at lag t equal to ρ^t, and rescaling of the time axis by a factor between 100 and 200, yielded a BIC value around -1010.900.

We now discuss a more complicated mixed modelling approach in section 3.2, and return to multivariate modelling with random effects in section 5.

3.2 A Semiparametric Additive Linear Mixed Model

Due to the limitations of the linear regression model described above, specifically the inflexible form of the linear predictor that quantifies the dose effect, we now fit a semiparametric additive linear mixed (SPALM) model of the form

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

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

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

where the matrix X contains the fixed effects predictors, matrix Z is the (basis function) design matrix in the semiparametric representation of the function of $f_1, \ldots, f_K$. We give brief details in the following sections.

3.2.1 Inference for the Linear Mixed Model

Inference for the linear mixed model in equation (3) is achieved using penalized least-squares [1], or likelihood procedures under a (model-based) assumption of Gaussianity; 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 $Y|\beta, u \sim N(X\beta + Zu, R)$, $u \sim N(0, G)$,

[1] The results are also justified under the (model-free) paradigm of Best Linear Unbiased Prediction (BLUP).

yielding (on integrating out u) the marginal model $Y|\beta \sim N(X\beta, ZGZ^{\mathsf{T}}+R)$. Let $V = ZGZ^{\mathsf{T}}+R$. Then the maximum penalized likelihood estimates of β and u given G and R are given by

$$\widehat{\theta} = \begin{bmatrix} \widehat{\beta} \\ \widehat{u} \end{bmatrix} = (C^{\mathsf{T}}R^{-1}C + B)^{-1}C^{\mathsf{T}}R^{-1}y \tag{4}$$

where $C = [X\ Z]$ and B is the block diagonal matrix with blocks $\mathbf{0}$ and G^{-1}. The variance of the estimators are given by $\mathrm{Var}[\widehat{\theta}] = (C^{\mathsf{T}}R^{-1}C + B)^{-1}$. Fitted values are obtained routinely as

$$\widehat{y} = C(C^{\mathsf{T}}R^{-1}C + B)^{-1}C^{\mathsf{T}}R^{-1}y$$

whereas predictions under this model at new design point $c_0 = [x_0\ z_0]$ are obtained from

$$\widehat{y}_0 = c_0(C^{\mathsf{T}}R^{-1}C + B)^{-1}C^{\mathsf{T}}R^{-1}y$$

with variance $c_0(C^{\mathsf{T}}R^{-1}C + B)^{-1}c_0^{\mathsf{T}}$. The quantities R and G that together define V and B can be estimated using maximum profile (integrated) likelihood

$$l_P(V) = \text{constant} - \frac{1}{2}\left[\log|V| + y^{\mathsf{T}}V^{-1}(I - X(X^{\mathsf{T}}V^{-1}X)^{-1}X^{\mathsf{T}}V^{-1})y\right]$$

obtained from the likelihood plugging in $\widehat{\beta} = (X^{\mathsf{T}}V^{-1}X)^{-1}X^{\mathsf{T}}V^{-1}y$, or REML, using the restricted likelihood

$$l_R(V) = l_P(V) - \frac{1}{2}\log|X^{\mathsf{T}}V^{-1}X|.$$

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 $\mathbb{R}^{\mathrm{ncol}(X)}$, 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 5.4, the fully Bayesian approach is described, and the results of using a diffuse specification for G are compared with results obtained when an informative prior specification is used.

3.2.2 Model Selection

Model selection can be carried out using BIC for linear mixed models, but an adjustment to the BIC calculation must be made. The conventional BIC is defined by

$$BIC = -2\hat{l} + (Number\ of\ Parameters \times \log(Sample\ size))$$

where $\hat{l}$ is the maximized log-likelihood. In a standard linear regression problem, the number of parameters fitted is given by the trace of the "hat" or smoother matrix, $\mathbb{S}$, that projects the data onto the fitted values, and that convention is followed here; we have

$$\widehat{y} = C(C^{\mathsf{T}}R^{-1}C + B)^{-1}C^{\mathsf{T}}R^{-1}y = \mathbb{S}y,$$

say, giving that the number of parameters is $\mathrm{tr}(\mathbb{S}) = \mathrm{tr}(C(C^{\mathsf{T}}R^{-1}C + B)^{-1}C^{\mathsf{T}}R^{-1})$.

3.2.3 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 section 5.4.3.

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

$$f_k(x) = \sum_{m=1}^{M} u_{km}(x - \kappa_{km})_+^q \tag{5}$$

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

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

In our analysis, we take $q = 1$, and use ten 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 (5) 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 in a straightforward fashion using `lme` in `R` for certain choices of the residual error covariance R, and more generally using numerical procedures for general covariance specifications. For the analysis in this paper we wrote code in `R` to perform the ML/REML estimation.

3.2.4 Results for the Absolute-Level data

In principal, many parts of the models of earlier sections can be replaced by semiparametric components. In the analysis presented here, for the absolute-level data, we fit $K = 3$ semiparametric components for the terms involving dose in the final models of sections 3.1.1 and 3.1.2, namely cumulative dose, D^c, and the interactions between D^c and age at interval, A, and D^c and time in occlusion t_O. We transform A by subtracting the minimum observed age, 36 months. For comparison, we also fit the model with the third semiparametric interaction excluded.

The results of the semiparametric analysis fit using empirical Bayes methods with ML estimation of the variance components are depicted in the left hand column of Figure 2. We report semiparametric functions with approximate 95% confidence bands, as inference for the fixed effects components was similar to that of the previous subsections. In Figure 2, the results from three models (0 - Fixed effects, 1 - Three semiparametric components, 3 - Two semiparametric components) are depicted. Figure 2 clearly shows the effect of cumulative dose, and the interaction between cumulative dose and age at interval, and the non-linear nature of these functions; note the difference between the estimated effects when only a fixed effect component is used. The BIC value corrected for the semiparametric modelling using the estimated number of parameters in section 3.2.2 is -1019.345 (uncorrected BIC is -933.682). This actually compares unfavourably with the

fixed-effect only model (BIC = -1057.044), although the log-likelihood values are comparable for both models.

3.3 Analysis of the Interval-Level data

We now analyze the vector of first differences of the visual acuity data, to study the change in visual acuity as a function of occlusion dose and the other covariates. Here, the dose in an interval, D, is used as a covariate, rather than cumulative dose, D^c, which was also tested but subsequently omitted. In addition to the variables defined above, let $Y_{ij} = Y^a_{ij} - V^a_{ij-1}$, for $j = 0, 1, ..., n_i,\ i = 1, 2, ..., N$, be the change visual acuity in interval j between visit $j-1$ and visit j for patient i. The variables, and interactions between them were fit using similar techniques to those used in section 3. Here, time on study, t, time in refraction, t_R, and time in occlusion, t_O, were tested as potentially influential covariates, but proved to add little to the fit of the model.

3.3.1 A Gaussian Linear Regression Model

Backwards step-wise Gaussian linear regression was implemented using BIC to discover the best fitting model. The models for the two phases that were found to be most appropriate were, for the Refraction phase, $L + P + T$, and for the Occlusion phase, $D + A + L + P + D.A + D.L + A.L$. Table 2 gives parameter estimates for the terms in the final models. The fit of this model yielded an estimate of the residual error standard deviation of $\sigma = 0.0901$, with an R^2 value of 0.470. When the regression models were fit separately to refraction and occlusion phase data, the estimates for σ were similar (0.1112 and 0.0788).

The Gaussian linear regression of 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 0.066 (0.036,0.096) between each visit, that strabismic children exhibit a similar improvement, but that children classified as Mixed type do not exhibit significant improvement. The stepwise procedure also reveals that visual acuity at the start of the phase of the study and visual acuity at the start of the interval are both influential in the visual improvement. For the occlusion phase, the beneficial effect of occlusion is apparent, but is modified by both the age of the child and the visual acuity at the start of the interval. The dose (in hours) main effect coefficient estimate is -7.630e-4 (3.778e-4). 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.

3.3.2 A Linear Mixed Effects 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 child grouping 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 a definition of this terminology). Autoregressive correlation structures were considered for the observed data, as in section 3.1.2; here, we retain the AR(1) model, where $\text{Corr}[Y_{ij}, Y_{ij'}] = \rho^{|j'-j|}$.

The BIC measure was used to select the most appropriate model, and 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, and that random slopes are not necessary to model the data. Table 2 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 $\sigma = 0.0735$ and $\rho = -0.1708$.

The parameter estimates from the mixed effect model are similar to those observed in the Gaussian linear regression. 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 0.083 (0.012,0.154) between each visit. Strabismic children exhibit a lesser degree of improvement while children of mixed type do not exhibit significant improvement. As in the Gaussian linear analysis, we see that 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.

3.3.3 A Semiparametric Additive Linear Mixed Model

A model similar to the one described in section 3.2 that uses a linear mixed model to give semi-parametric inference. For the interval-level data, we utilized three semiparametric components; 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 examined two covariance structures for the repeated measures - an exponential decay-in-time covariance and an AR(1) model in interval number. We used 10 fixed knots, with positions determined by covariate quantiles, but the results were robust to specifications with up to 50 knots. For the interval-level data, we assumed three different forms for the residual autocovariance for the repeated measures component - uncorrelated errors, exponential decay in autocorrelation in time (in days) between observations, and AR(1) autocorrelation in clinic visit number - but there was effectively no difference in the results for each in the resulting estimates of the semiparametric components.

The results of the semiparametric analysis fit using empirical Bayes methods with ML estimation depicted in the left column of Figure 3, which again shows the interaction between dose and age at interval and current visual acuity. The non-linear nature of the dose-response nature is captured by the semi-parametric model. Note that in this case, the effect of dose (outside of its interaction with the other covariates) is minimal. We retain dose in the model to maintain marginal coherence (a term fitted in an interaction cannot be omitted as a main effect).

3.4 Comments on the Regression and SPALM analyses

The analyses in section 3 has confirmed the beneficial effect of dose on vision in children in the study base. Furthermore, we have quantified the dose-response relationship, and identified the influential interactions between the covariates. Overall, the broad pattern of results 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. We have also estimated the standard deviation of the residual errors to be of the order of 0.1 (across both analyses); this level of residual error is rather higher than that predicted by ophthalmological opinion.

A potential problem with the analysis 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. On the basis of the analysis presented, we perhaps cannot truly ascribe a *causal* effect to dose on the improvement in vision. We address this issue in the following section.

4 Causal Approaches to Estimating the Dose-Response Relationship

A potential problem with approaches used in section 3 is that they do not recognize the lack of randomization in the amount occlusion dose received, raising the possibility that the effect of occlusion treatment is misrepresented by the estimates presented. We wish to ascertain the true effect of dose, and thus a causal analysis which accounts for the potential confounding between dose levels and other covariates may be necessary. We present such an analysis in this section. The principal tool that we use 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.

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

$$r(d, x) = f_{D|X}(d|x) \tag{6}$$

that is, the conditional density function for D given that $X = x$, evaluated at $D = d$, and the corresponding random quantity, $R = r(D, X)$. The GPS is the extension of the propensity score defined by Rosenbaum and Rubin (1983) to continuous treatments.

When applied appropriately, the GPS has a balancing property, in that within score strata the conditional density value for $D = d$ does not depend on X, which is a checkable assumption for any proposed scoring procedure. 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; we first build a (regression) model to predict D given X that yields R, and then a second (regression) model for Y given D and R that has a built-in balancing of any possible confounding between D and X that may bias our estimate of the effect of D on Y.

4.2 The Average Potential Outcome

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

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

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 procedure for doing the estimation is given in section 4.3, and depends on plug-in prediction of the dose effect. In section 5.5, we demonstrate how a simple Bayesian procedure can be used to propagate uncertainty through the model in a coherent fashion. In the discussion below, for illustration we restrict attention to the occlusion phase data only, and focus first on the interval-level data, but implementation of the methods for the remaining data is straightforward.

4.3 Applying the GPS to the MOTAS data

The approach to estimating the APO at dose level d, $\mu(d)$, outlined by Hirano and Imbens (2004), proceeds as follows:

I. Estimate β in the predictive model for D given $X = x$, $f_{D|X}(d|x, \beta)$.

II. Compute the estimated GPS, $\widehat{r}_i = f_{D|X}(d_i|x_i, \widehat{\beta})$.

III. Estimate α in the predictive model for Y given $D = d$ and $R = \widehat{r}$, $f_{Y|D,R}(y|d, \widehat{r}, \alpha)$.

IV. Estimate the APO at dose level d by

$$\widehat{\mu}(d) = \widehat{E}[Y|D = d] = \frac{1}{N}\sum_{i=1}^{N} E_{Y|D,R}[Y_i(d)|D = d, \widehat{r}_i = \widehat{r}(d, x_i), \widehat{\alpha}]$$

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

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 $\xi(d, x)$.

Several components in this model must be user-specified; the two key conditional models $f_{D|X}(d|x, \beta)$ and $f_{Y|D,R}(y|d, r, \alpha)$ must be selected to reflect the various relationships between the variables. However, the adequacy of both components is checkable in a straightforward statistical fashion.

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. This suggests that the predictive model $f_{D|X}(d|x, \beta)$ should 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 if we propose a parametric distribution 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.

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 for D_+, it was discovered that interval number, visual acuity at start of interval, L, length of interval (in days), and amblyopic type, T, all were 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 GPS

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

where

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

For the GPS to be effective, we require that within strata of $\hat{r}$, the distribution of D should not depend on X. A check of the distribution within quintile categories of $\hat{r}$ reveals no apparent relationship between D and X (see, for illustration, Figure 4).

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 estimated GPS score, $\widehat{R}$. In Hirano and Imbens (2004), only the expected value of Y given D and R is considered; we follow that strategy here, modelling

$$E_{Y|D,R}[Y|D = d, R = r(d,X), \alpha] = \alpha_0 + \alpha_1 d + \alpha_2 r + \alpha_{12} d.r \tag{7}$$

and estimate the parameters using least-squares, although a model assuming independent Gaussian residual errors does appear sensible, and facilitates a likelihood analysis. Using the model obtained above, 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, using the empirical average over the 411 interval observations as in IV of the strategy above, we obtain an estimate of the average potential effect dose/response under the specified model.

4.3.3 Results

A plot of the dose/response curve is presented in Figure 5. This plot indicates that the causal effect of dose, 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 (that is, corresponds to vision improvement) over the entire range of positive doses considered. A numerical summary for the APO for the frequentist method is found in the first part of Table 5. Comparing the results of this analysis with the SPALM analysis for the interval-level data (specifically, comparing Figure 3 with Figure 5, we note that the estimated causal effect of dose is

somewhat less than that reported in the SPALM analysis, although the two results are difficult to compare directly.

The model used, for example, in equation (7) can be readily extended to a more flexible model; here, we checked the qualitative changes in the inferences made after the inclusion of quadratic terms in the linear predictor, but they were minimal.

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 interactions between the covariates render the selected model difficult to interpret. Therefore, to study the impact of different dosing strategies, we now implement a Bayesian analysis, as this propagates the uncertainty in the inference in a fully coherent fashion.

In the case of the fixed effects model, under the assumption of independent residual errors, computation of the posterior quantities of interest can proceed analytically in both the absolute-level and interval-level data. For brevity, we focus on the correlated response regression model of section 3.1.1 for absolute-level data with and without random effects, and perform inference using Markov chain Monte Carlo, but note that any of the models fit in earlier sections could be implemented in a Bayesian MCMC setting.

5.1 Bayesian Posterior Calculation for Repeated Measures Data

Consider a linear model formulation using the notation introduced earlier, that is where $Y \sim N(X\beta, R)$, where R is a block diagonal error covariance matrix $R = \text{diag}(R_1, \ldots, R_N)$. For example, the components of R can be specified via the exponential decay, or AR(1) autocorrelation functions. We focus on the former for illustration.

To complete the specification, we use a diffuse (improper uniform) prior specification for β, and an improper Jeffreys-type prior on the positive parameters in the exponential autocovariance function, that is, we take $p(\beta, \lambda, \zeta, \nu) = (\lambda\zeta\nu)^{-1}$, and derive the posterior distribution. This factorizes $p(\beta, \lambda, \zeta, \nu|y) = p(\lambda, \zeta, \nu|y)\, p(\beta|y, \lambda, \zeta, \nu)$ where

$$p(\lambda, \zeta, \nu|y) \propto \frac{|\ M_3|^{-1/2}}{\prod_{i=1}^{N} |R_i|^{1/2}} \exp\left\{-\frac{1}{2}\left[M_1 - M_2^{\mathsf{T}} M_3^{-1} M_2\right]\right\} \frac{1}{\lambda\zeta\nu} \tag{8}$$

where

$$M_1 = \sum_{i=1}^{N} y_i^{\mathsf{T}} R_i^{-1} y_i \qquad M_2 = \sum_{i=1}^{N} X_i^{\mathsf{T}} R_i^{-1} y_i \qquad M_3 = \sum_{i=1}^{N} X_i^{\mathsf{T}} R_i^{-1} X_i$$

and $\beta|y, \lambda, \zeta, \nu \sim N_p(M_3^{-1} M_2, M_3)$. The posterior distribution in equation (8) is not available analytically, but inference may be carried out using Markov chain Monte Carlo (MCMC) on the three parameter joint posterior. We use a Metropolis update on a sweep of the conditionals, reparameterized onto the log scale, and jointly on the block of the three parameters. The conditional posterior for β given (λ, ζ, ν) can be sampled directly.

A fully Bayesian analysis is also possible for the SPALM model of section 3.2. In the notation of that section, the posterior distribution of the covariance parameters is identical to that in (8), but with

$$M_1 = \sum_{i=1}^{N} y_i^{\mathsf{T}} R_i^{-1} y_i \qquad M_2 = \sum_{i=1}^{N} C_i^{\mathsf{T}} R_i^{-1} y_i \qquad M_3 = \sum_{i=1}^{N} C_i^{\mathsf{T}} R_i^{-1} C_i = C^{\mathsf{T}} R^{-1} C.$$

where $R \equiv R(\lambda, \zeta, \nu)$, and the posterior distribution for $\theta = [\beta\ u]^{\mathsf{T}}$ is multivariate normal (dimension $p + MK$) with mean and variance

$$\mu = (C^{\mathsf{T}} R^{-1} C + B)^{-1} C^{\mathsf{T}} R^{-1} y \qquad \Sigma = (C^{\mathsf{T}} R^{-1} C + B)^{-1}$$

respectively. In addition, rather than using an improper uniform prior for β, an informative prior can also be specified. In this case, the calculation proceeds in an identical fashion, with the marginal posterior for (λ, ζ, ν) sampled using Metropolis-Hastings, and the conditional posterior for β (or (β, u)) multivariate normal.

To extend the mixed model, a further level can be added to the hierarchy in some cases, although this is not sensible for the semiparametric components. For example, fitting a random effects model similar to that in equation (1) is straightforward using a Gibbs sampler. Denoting by $\boldsymbol{\eta} = (\eta_1, ..., \eta_N)$ the vector of child-specific random effects (intercepts), the posterior of interest becomes the joint distribution $p\left(\theta, \lambda, \zeta, \nu, \boldsymbol{\eta}, \sigma_\eta^2 | y\right)$, where σ_η^2 is the (unknown) random effect error variance, which is included in the MCMC cycle; we might assign an Inverse Gamma prior with parameters 2.5 and 0.25. Then, conditional on η, the posterior for $(\theta, \lambda, \zeta, \nu)$ is updated as in the fixed effect only model, with datum y_{ij} replaced by $y_{ij} - \eta_i$. Conditional on $(\theta, \lambda, \zeta, \nu)$ and σ_η^2, the posterior for η_i is univariate Gaussian. Finally, conditional on all other parameters, the posterior for σ_η^2 is Inverse Gamma.

5.2 Bayesian Inference for the Absolute-Level Data

For the absolute-level data, for the covariate model, to specify the design matrices, we fit the model from section 3.1.1. MCMC code was written in `R`, and in version 2.0.1 took around 10 minutes to implement on a 1.6GHz machine. After appropriate construction of an independence Metropolis kernel for the parameters in the autocovariance model using short pilot runs, it was possible to produce samples from the marginal posterior distribution for parameters $(\beta, \lambda, \zeta, \nu)$ that were virtually uncorrelated serially. The numerical summaries of the marginal posterior distribution for each of the parameters are given in Table 3.

The beneficial effect of an accumulating occlusion dose is evidenced by the negative coefficient of D^c, with a 95% credible interval of (-1.770e-03,-8.825e-04). However, this effect is moderated by the age at interval of the child, A, and also the time the child has spent in occlusion, t_O. The sample median and 95% posterior credible intervals for the residual error standard deviation in the two analyses were similar; 0.147 (0.135,0.161) for fixed effects model, 0.135 (0.121,0.155) for the random effects. The posterior samples from the fixed-effect analysis will be used in section 5.6 to examine the impact of different dosing strategies.

5.3 Bayesian Inference for the Interval-Level Data

A Bayesian analysis of the interval-level data is again facilitated using MCMC for the multivariate response data, with or without random effects. For brevity, we report only the results of the fized

effects analysis; these are contained in Table 4. The results of earlier sections are confirmed - there appears to be a significant beneficial effect of dose (95% credible interval (-1.55e-03,-4.28e-05), but that effect is moderated by the age of the child at the start of the interval. Furthermore, the worse the vision of the child at the start of the interval, the greater degree of improvement in vision (95% credible interval (-1.56e-03,-1.90e-04) for the coefficient of the $D.L$ interaction).

5.4 Bayesian Semiparametric Modelling

5.4.1 Analysis of the Absolute Level Data

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 5.1 and equation (8). 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 (described in section 5.4.3 below) are depicted in Figure 2. We omit details of the posterior summaries and focus only on the semiparametric components.

5.4.2 Analysis of the Interval Level Data

A semiparametric model similar to the one described in section 3.3.3 can be fitted in the Bayesian framework. Again, 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 are depicted in Figure 3.

5.4.3 An Informative Prior Specification for the SPALM

In the SPALM model, the specification of random effects prior matrix G can be engineered to match prior opinion about the nature (that is, smoothness or curvature) of the modelled function. Consider a single semiparametric component $Y = Zu$, where $u \sim N(0, G)$, so that $Y \sim N(0, ZGZ^{\mathsf{T}})$, where we require that *a priori* $Y \sim N(0, V_0)$. Then

$$V_0 = ZGZ^{\mathsf{T}} \Longrightarrow G = (Z^{\mathsf{T}}Z)^{-1}Z^{\mathsf{T}}V_0Z(Z^{\mathsf{T}}Z)^{-1}$$

and G should adopt a data-dependent form, giving a prior that is similar in structure to the "g-prior" (Zellner (1983)). Conditional on knot points $\kappa_1, \ldots, \kappa_M$, we can specify any required prior autocovariance structure. For example, we could specify a prior with high autocorrelation, thereby encouraging smoothness in the semiparametric component. In our analysis, we specify V_0 to be a diagonal matrix such that the prior variation in the semiparametric function is concentrated on the

range ±2. This results in a required prior variance for the dose component to be specified by

$$G^{-1} = 10^{-3} \begin{pmatrix} 14 & 13 & 12 & 11 & 11 & 10 & 9 & 8 & 6 & 5 & 2 \\ 13 & 12 & 11 & 11 & 10 & 9 & 8 & 7 & 6 & 4 & 2 \\ 12 & 11 & 11 & 10 & 9 & 9 & 8 & 7 & 6 & 4 & 2 \\ 11 & 11 & 10 & 10 & 9 & 8 & 8 & 7 & 6 & 4 & 2 \\ 11 & 10 & 9 & 9 & 9 & 8 & 7 & 6 & 5 & 4 & 2 \\ 10 & 9 & 9 & 8 & 8 & 7 & 7 & 6 & 5 & 4 & 1 \\ 9 & 8 & 8 & 8 & 7 & 7 & 6 & 5 & 5 & 3 & 1 \\ 8 & 7 & 7 & 7 & 6 & 6 & 5 & 5 & 4 & 3 & 1 \\ 6 & 6 & 6 & 6 & 5 & 5 & 5 & 4 & 4 & 3 & 1 \\ 5 & 4 & 4 & 4 & 4 & 4 & 3 & 3 & 3 & 2 & 1 \\ 2 & 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

We note here three things. First, this is a much more precise specification than the noninformative prior we selected. Secondly, it is much less precise than the prior deduced using the empirical Bayes procedure based on ML/REML estimation of the components of G; the parameters, $(\widehat{\sigma}_1, \widehat{\sigma}_2, \widehat{\sigma}_3)$, that define the diagonal components of G for the absolute-level data are estimated as (2.053,0.009,0.140), and for the interval-level data are (7.589e-08, 2.314e-06,7.430e-04) respectively. Finally, as the prior is design-dependent, this specification is only strictly appropriate for the fixed-knot case, and will change in a straightforward fashion when the knot positions change. However in section 5.4.4 we also note the results of an analysis where the knots are allowed to move in the MCMC analysis.

The results from an analysis using this informative prior (and fixed knots) are depicted in Figures 2 and 3 (right columns), where results for the Non-Informative and Empirical Bayes Priors are also depicted for comparison. Overall, the results for absolute- and interval-level data are broadly similar when the two fully Bayesian procedures are used, but the magnitude of the various dose effects are estimated to be much larger than those estimated using the empirical Bayes procedures. We note that the deduced empirical Bayes prior has **extremely** (we argue unreasonably) high precision for several of the components, and prefer the informative specification.

5.4.4 Relaxing the assumption of fixed knot positions

As a final check of the reported effect magnitude, we relax the assumption that knot positions in the truncated basis spline SPALM analysis of previous sections are fixed. We treat the knot positions as further unknown parameters to be estimated, and assign a prior distribution that states that the knot positions are the order statistics from a Uniform sample on the range of the variable concerned. Incorporating this into the MCMC scheme is straightforward; we augment the algorithm described in section 5.1 with a further Metropolis-Hastings step that selects a knot at random, samples a candidate new knot position from its conditional prior distribution, and performs the usual accept/reject step.

The results from the MCMC analysis comparing fixed and movable knots for the absolute- and interval-level data (using the informative prior from section 5.4.3, and an exponential residual error autocorrelation) are depicted in Figure 6. In the right hand column, for the interval-level data, the main difference in the results is that the 95% pointwise credible intervals are slightly wider for the movable knots analysis. In the left hand column, however, the Bayesian posterior

estimated semiparametric response functions include features in the movable knots analysis that are not present in the fixed knots analysis, especially for small values of the functions for cumulative dose D^c and the interaction between D^c and translated age at interval, $A - 36$. Inspection of the data reveals that this extra "kink" corresponds to a pair of outlying observations that occur in adjacent intervals for one child in the study. The influence of these points is being smoothed away by the other analyses.

5.5 Bayesian Causal Analysis: The Bayesian Propensity Score

The causal 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/causal 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 average potential outcome 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 as follows.

5.5.1 MCMC for the Bayesian Propensity Score

To sample the posterior distribution for (α, β, γ), given the data (y, d, x), iterate around the following cycle with parameter updating: at iteration m, let the current values of the parameters be $(\alpha^{(m)}, \beta^{(m)}, \gamma^{(m)})$, and let $r_1^{(m)}, \ldots, r_N^{(m)}$ be defined by

$$r_i^{(m)} = f_{D|X}(d_i|x_i, \beta^{(m)}, \gamma^{(m)}).$$

Let $p(\alpha, \beta, \gamma)$ be the joint prior distribution for the three parameters, and let $p(\alpha)$ and $p(\beta, \gamma)$ be the corresponding marginal priors. Then

1. Sample $\alpha^{(m+1)}$ from full conditional $p(\alpha|y, d, x, \beta^{(m)}, \gamma^{(m)})$ using Metropolis-Hastings, where

   $$p(\alpha|y, d, x, \beta, \gamma) \propto f_{Y|D,R}(y|d, r, \alpha)p(\alpha)$$

   proposing from some appropriate distribution and accepting/rejecting in the usual fashion.

2. Sample $(\beta^{(m+1)}, \gamma^{(m+1)})$ from full conditional $p(\beta, \gamma|y, d, x, \alpha^{(m+1)})$ using Metropolis-Hastings as follows

   (i) Propose candidate values $(\beta^{(new)}, \gamma^{(new)})$ from some appropriate distribution, q, possibly functionally dependent on the current values $(\beta^{(m)}, \gamma^{(m)})$.

   (ii) Compute for $(\beta^{(new)}, \gamma^{(new)})$

   $$r_i^{(new)} = f_{D|X}(d_i|x_i, \beta^{(new)}, \gamma^{(new)}).$$

(iii) For brevity, let

$$L(\beta, \gamma; \alpha) = f_{Y|D,R}(y|d, r, \alpha) f_{D|X}(d|x, \beta, \gamma)$$

where the first term depends on β and γ through r. Define Λ as the minimum of 1 and

$$\frac{L(\beta^{(new}\gamma^{(new)}; \alpha^{(m+1)})}{L(\beta^{(m)}, \gamma^{(m)}; \alpha^{(m+1)})} \frac{p(\beta^{(new)}, \gamma^{(new)})}{p(\beta^{(m)}, \gamma^{(m)})} \frac{q(\beta^{(m)}, \gamma^{(m)}|\beta^{(new)}, \gamma^{(new)})}{q(\beta^{(new)}, \gamma^{(new)}|\beta^{(m)}, \gamma^{(m)})}.$$

where the L terms depend on $r^{(new)}$ and $r^{(m)}$) numerator and denominator respectively.

(iv) With probability Λ, set new values for parameters $(\beta^{(m+1)}, \gamma^{(m+1)})$ equal to $(\beta^{(new)}, \gamma^{(new)})$, else set them equal to $(\beta^{(m)}, \gamma^{(m)})$.

After a sufficient number of iterations of this scheme, the required sample from the joint posterior distribution is obtained.

5.5.2 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 (hypothetical) 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 4.3, 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. Thus 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 CPO. In fact, the Bayesian approach yields the entire posterior distribution for the CPO, which can be summarized via quantile summaries for, say, 95% credible intervals.

5.5.3 Results

We restrict attention to the interval-level data. 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 (7) as the model for mean effect. The results are depicted in Figure 7, with the purely frequentist estimated APO for varying dose D from section 4.3 included for comparison. A numerical summary is provided in the second half of Table 5. In this example, there appears to be little qualitative difference between the frequentist and Bayesian inferences.

5.6 Bayesian Prediction: The Impact of Different Dosing Strategies

The Bayesian analysis of the previous sections is useful as it readily facilitates a study of the impact of different dosing strategies. Once a large sample from the posterior distribution has been obtained, a sample from the posterior predictive distribution can be obtained for fixed values of the covariates by sampling from the likelihood part of the model, given on each posterior sample in turn.

We examine the impact of different dosing strategies on two hypothetical children, one aged 48 months at the start of study, the other aged 60 months. Each child will enter the study with a visual acuity of either 1.0 or 0.6 logMAR, and be diagnosed as of anisometropic type. They will spend four months in refraction, then be switched to occlusion. In occlusion, they could receive 2, 6 and 9 hours of occlusion per day, over a 30 day month period, and then be followed for a further six visits, when the study will end.

The impact of the different dosing strategies for the two children is depicted in the four panels of Figure 8, where the posterior predictive median response profiles are plotted, where the posterior samples are generated in the Bayesian analysis of the absolute-level data in section 5.2. The difference between the response profiles under different strategies is evident (the expected profile under a zero dose strategy is included for comparison). Also, comparing the results for the two children, the impact of age is striking, with the younger child improving to a greater degree than the older child under each strategy. This analysis is illustrative; the impact for children with varying characteristics can readily be studied in the same way.

6 Discussion

The MOTAS study has facilitated, for the first time, the efficacy of occlusion therapy in the treatment of amblyopia to be quantified. In this paper, we quantified the improvement in visual acuity of amblyopic children in refractive adaption, as well as the dose-response relationship between improvement in vision and occlusion.

We have inspected the data in two ways. For the absolute-level data, key determinants of the change in visual acuity amongst the measured variables were identified. It was evident that occlusion dose was an influential factor and, using a semiparametric mixed effect model, we were able to identify and quantify the marginal dose-response impact of dose on visual acuity. For the interval-level data, the relationship between visual acuity and occlusion dose in an interval was modelled using similar regression methods, and the impact of different levels of occlusion dose per interval was studied.

Our analysis demonstrates that the relationship between dosing and age is complex. We found occlusion to be successful in children up to eight years of age, however the same amount of occlusion was less effective in children who were older or had worse visual acuity at baseline than younger or better-seeing children. We have provided further evidence that age at treatment does influence the effectiveness of occlusion; this may contribute to the current debate on the best age(s) at which to conduct childhood vision screening.

Finally, we studied models and procedures that allowed study of the causal effect of occlusion dose. For this analysis we used a propensity score approach that constructed a variable, the Generalized Propensity Score, on which to regress (in conjunction with dose) the by-interval improvement response. Using this approach, which aims to adjust for confounding between the regressors, we

obtained estimates of the average (potential) response to different levels of dosing.

We now have a framework to examine the impact of different occlusion dosing strategies. This will enable clinicians to discuss the time scales for the occlusion component of therapy with parents. An area for future research is the development of an optimal dynamic regime. Dynamic regimes allow treatment to change over time based on patient history - for example, an optometrist may prescribe a tapering of occlusion as improvements in visual acuity are recorded over time. Dynamic regimes are more like clinical care than traditional clinical studies and studies incorporating such regimes may see higher concordance with treatment. A number of innovations in the design and analysis of dynamic regime trials have recently been made (see, for example, Lavori and Dawson (2004); Murphy (2004); Robins (2004)); an implementation in relation to the amblyopia study can be found in Moodie et al. (2005).

Acknowledgement: The first author was supported by Merck Graduate Fellowship. The MOTAS study was supported by The Guide Dogs for the Blind Association, UK.

(Chapter head:)*

Bibliography

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

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

Hall, D. M. (1989). *Health for all children* (4 ed.)., Chapter Screening for vision defects, pp. 230–244. Oxford: Oxford Medical Publications.

Hirano, K. and G. W. Imbens (2004). The propensity score with continuous treatments. Technical report, University of Miami.

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

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

Moodie, E. E. M., T. S. Richardson, and D. A. Stephens (2005). Demystifying optimal dynamic treatment regimes. UW Biostatistics Working Paper 240, University of Washington, http://www.bepress.com/uwbiostat/paper240.

Murphy, S. A. (2004). An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine*. In Press.

PEDIG (2003). A randomized trial of patching regimens for treatment of moderate amblyopia in children. *Archives of Ophthalmology 121*(5), 603–611.

Rahi, J. S., S. Logan, C. Timms, I. Russell-Eggitt, and D. Taylor (2002). Risk, causes and outcomes of visual impairment after loss of vision in the non-amblyopic eye: a population-based study. *Lancet 360*(2), 597–602.

Repka, M. X., D. K. Wallace, R. W. Beck, R. T. Kraker, E. E. Birch, S. A. Cotter, S. Donahue, D. F. Everett, R. W. Hertle, J. M. Holmes, G. E. Quinn, M. M. Scheiman, and D. R. Weakley (2005). Two-year follow-up of a 6-month randomized trial of atropine vs patching for treatment of moderate amblyopia in children. *Archives of Ophthalmology. 123*(2), 149–157. On behalf of the Pediatric Eye Disease Investigator Group (PEDIG).

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

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

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

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

Stewart, C. E., M. J. Moseley, A. R. Fielder, and D. A. Stephens (2004). Optimization of the dose-response of occlusion therapy for amblyopia: the ROTAS study. *Investigative Ophthalmology and Visual Science (suppl.) 45*, 2579.

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

Zellner, A. (1983). Applications of Bayesian analysis and econometrics. *The Statistician 132*, 23–34.

Table 1: Estimates from the ML and REML model fit to absolute-level data of the multivariate Gaussian model from section 3.1.1; t.stat is the Wald test statistic, p is the corresponding p-value.

		ML				REML			
Phase	Term	Est.	s.e.	t.stat	p	Est.	s.e.	t.stat	p
Ref.	Int.	1.569e-2	2.99e-2	0.524	0.600	1.638e-2	3.083e-2	0.531	0.595
	t_R	-1.041e-3	1.209e-4	-10.112	0.000	-1.046e-3	1.049e-4	-9.969	0.000
	t_O	8.690e-5	2.477e-4	0.350	0.726	8.363e-5	2.516e-4	0.332	0.740
	T_M	5.150e-2	3.117e-2	1.652	0.098	5.115e-2	3.231e-2	1.583	0.113
	T_S	6.723e-3	3.177e-2	0.305	0.760	6.545e-3	3.212e-2	0.203	0.839
	S	8.744e-1	3.654e-2	23.927	0.000	8.746e-1	3.768e-2	23.214	0.000
Occ.	Occ.	-1.593e-1	5.569e-2	-2.862	0.004	-1.590e-1	5.617e-2	-2.830	0.005
	S	-6.450e-2	3.102e-2	-2.079	0.038	-6.362e-1	3.124e-2	-2.037	0.041
	A	1.729e-3	7.005e-4	2.469	0.014	1.720e-3	7.068e-4	2.434	0.015
	D^c	-1.360e-3	2.138e-4	-6.361	0.000	-1.363e-3	2.171e-4	-6.280	0.000
	$D^c.A$	1.286e-5	3.110e-6	4.139	0.000	1.286e-5	3.150e-6	4.077	0.000
	$D^c.t_O$	1.933e-6	3.431e-7	5.674	0.000	1.940e-6	3.524e-7	5.611	0.000
	$t_O.S$	-1.734e-3	2.870e-4	-6.041	0.000	-1.734e-3	2.922e-4	-5.936	0.000
Covariance Parameters		$\widehat{\lambda}$ =0.020 $\widehat{\zeta}$ = 158.856 $\widehat{\nu}$ = 0.617				$\widehat{\lambda}$ =0.021 $\widehat{\zeta}$ =169.746 $\widehat{\nu}$ = 0.619			

Table 2: Estimates and standard errors for the parameters from a Normal linear model and from a random intercepts mixed effects model with AR correlation for the interval level data.

		Model with Fixed Effects Only				Model with Random Effects			
Phase	Term	Est.	s.e.	t.stat	p	Est.	s.e.	t.stat	p
Refraction	Int.	-6.573e-2	1.539e-2	-4.271	0.000	-8.509e-2	3.149e-2	-2.702	0.007
	L	-3.571e-1	4.231e-2	-8.441	0.000	-6.376e-1	4.145e-2	-15.383	0.000
	P	2.358e-1	4.330e-2	5.445	0.000	4.839e-1	5.405e-2	8.952	0.000
	T_M	6.066e-2	1.724e-2	3.518	0.000	8.313e-2	3.640e-2	2.284	0.023
	T_S	6.569e-3	1.755e-2	0.374	0.708	1.682e-2	3.818e-2	0.441	0.660
Occlusion	Int.	-1.229e-3	3.500e-2	-0.035	0.972	-5.728e-3	3.473e-2	-0.165	0.869
	D	-7.630e-4	3.778e-4	-2.019	0.044	-8.645e-4	3.226e-4	-2.680	0.008
	A	-6.041e-4	5.121e-4	-1.180	0.239	-5.738e-4	4.920e-4	-1.166	0.244
	L	-5.231e-1	6.209e-2	-8.425	0.000	-5.113e-1	5.557e-2	-9.237	0.000
	P	1.146e-1	2.209e-2	5.189	0.000	1.234e-1	2.144e-2	5.755	0.000
	$D.A$	1.113e-5	5.930e-6	1.878	0.061	1.339e-5	5.063e-6	2.645	0.008
	$D.L$	5.930e-6	3.472e-4	-2.586	0.010	-9.705e-4	2.959e-4	-3.279	0.001
	$A.L$	5.247e-3	9.724e-4	5.396	0.000	4.912e-3	9.242e-4	5.315	0.000

Table 3: Posterior Summaries for the parameters in the Bayesian analysis of the fixed effect model for the absolute-level data. Summaries derived from 5000 posterior samples.

					Quantiles		
	Phase	Term	Mean	S.D.	0.025	0.500	0.975
Fixed Effects	Ref.	Int.	1.537e-02	3.152e-02	-4.580e-02	1.536e-02	7.771e-02
		t_R	-1.044e-03	1.060e-04	-1.255e-03	-1.044e-03	-8.365e-04
		t_O	8.148e-05	2.517e-04	-3.976e-04	8.239e-05	5.819e-04
		T_M	5.275e-02	3.300e-02	-9.924e-03	5.250e-02	1.174e-01
		T_S	7.407e-03	3.244e-02	-5.841e-02	8.069e-03	7.076e-02
		S	8.742e-01	3.800e-02	7.985e-01	8.746e-01	9.488e-01
	Occ.	Occ.	-1.604e-01	5.665e-02	-2.712e-01	-1.611e-01	-4.853e-02
		S	-6.405e-02	3.147e-02	-1.242e-01	-6.454e-02	-1.750e-03
		A	1.748e-03	7.156e-04	3.293e-04	1.755e-03	3.163e-03
		D^c	-1.354e-03	2.195e-04	-1.791e-03	-1.351e-03	-9.302e-04
		$D^c.A$	1.273e-05	3.167e-06	6.601e-06	1.267e-05	1.910e-05
		$D^c.t_O$	1.939e-06	3.517e-07	1.265e-06	1.935e-06	2.628e-06
		$t_O.S$	-1.729e-03	2.888e-04	-2.314e-03	-1.725e-03	-1.171e-03

Table 4: Posterior Summaries for the parameters in the Bayesian analysis of the fixed effect model for the interval level data: 5000 posterior samples.

				Quantiles		
Phase	Term	Mean	S.D.	0.025	0.500	0.975
Refraction	Int.	-6.69e-02	1.64e-02	-9.90e-02	-6.69e-02	-3.54e-02
	T_M	6.43e-02	1.83e-02	3.01e-02	6.38e-02	1.01e-01
	T_S	4.62e-03	1.88e-02	-3.21e-02	4.71e-03	4.18e-02
	L	-4.02e-01	5.56e-02	-5.09e-01	-4.02e-01	-2.94e-01
	P	2.73e-01	5.33e-02	1.68e-01	2.73e-01	3.77e-01
Occlusion	Int.	-1.18e-02	3.86e-02	-8.93e-02	-1.20e-02	6.40e-02
	P	1.31e-01	2.66e-02	8.09e-02	1.30e-01	1.85e-01
	L	-5.35e-01	6.76e-02	-6.67e-01	-5.34e-01	-4.03e-01
	A	-4.83e-04	5.61e-04	-1.59e-03	-4.85e-04	6.33e-04
	D	-7.83e-04	3.86e-04	-1.55e-03	-7.82e-04	-4.28e-05
	$D.A$	1.14e-05	6.11e-06	-6.24e-07	1.14e-05	2.33e-05
	$D.L$	-8.91e-04	3.52e-04	-1.56e-03	-8.94e-04	-1.90e-04
	$A.L$	5.09e-03	1.05e-03	2.99e-03	5.10e-03	7.13e-03

Table 5: 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

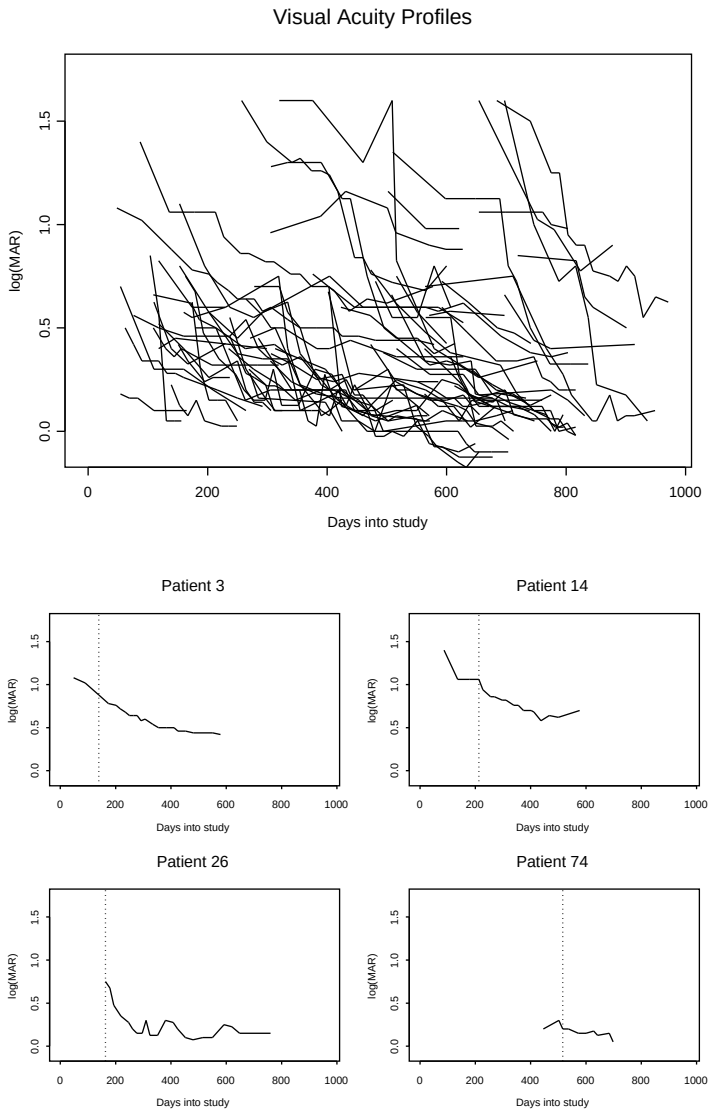

Figure 1: Profile plots for the individuals in the MOTAS study (top) and for four selected patients, with the start of occlusion indicated by the dotted line (bottom).

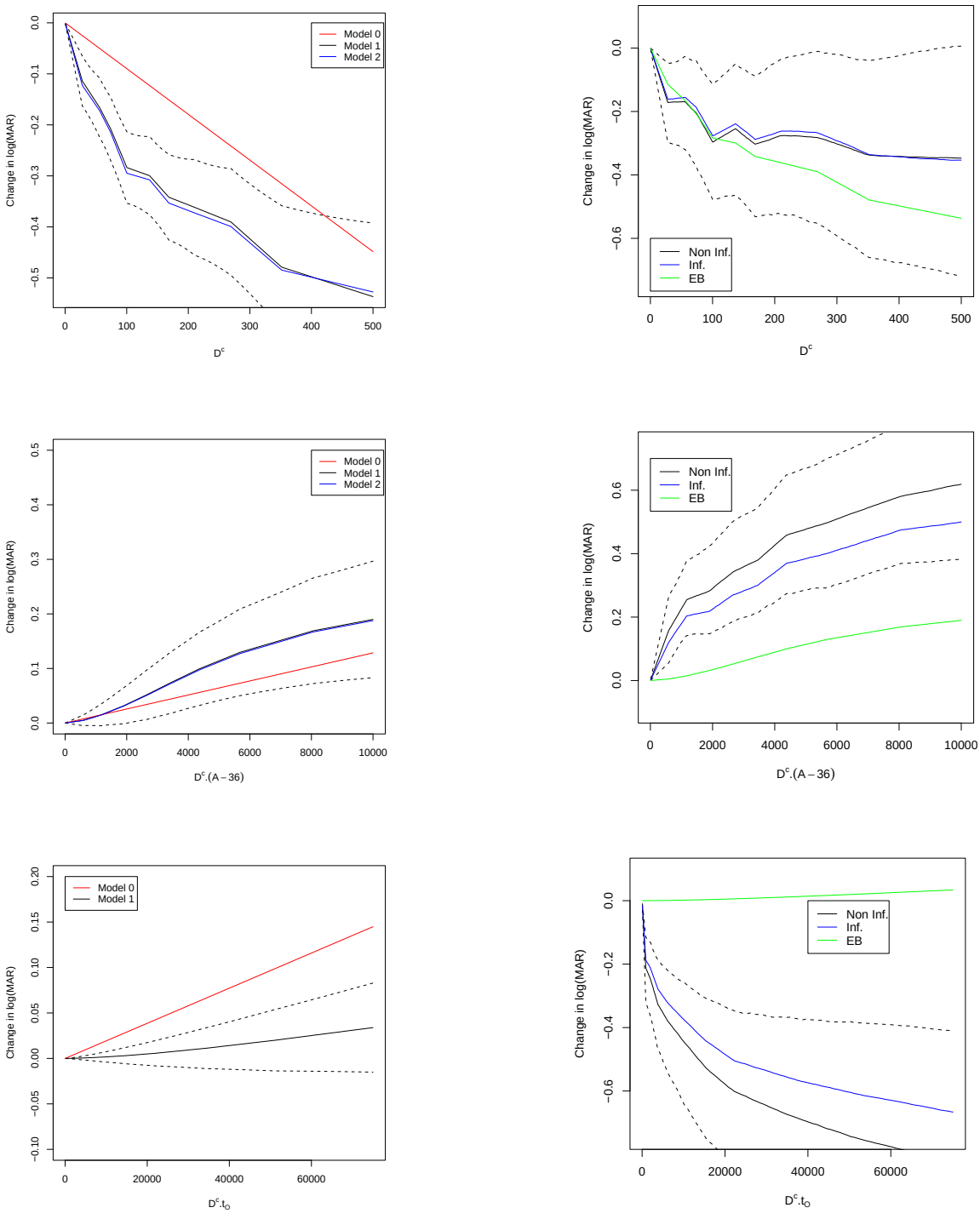

Figure 2: Estimated semiparametric functions from the linear model and SPALM analyses of the absolute-level data from section 3.2. Top panel for cumulative dose D^c, middle panel for the interaction between D^c and $A - 36$, bottom panel for the interaction between D^c and t_O. Left column for maximum likelihood (empirical Bayes) estimates, right column for fully Bayesian Model using Non-Informative (black) and Informative (blue) prior specification for the random effects coefficients. Empirical Bayes estimates (green) included for comparison. Dotted lines are 95% pointwise credible intervals from the Non-Informative prior model fit. Model 0 is Fixed effects only, Model 1 uses three semiparametric components, Model 2 uses two semiparametric components. Some lines overlaid.

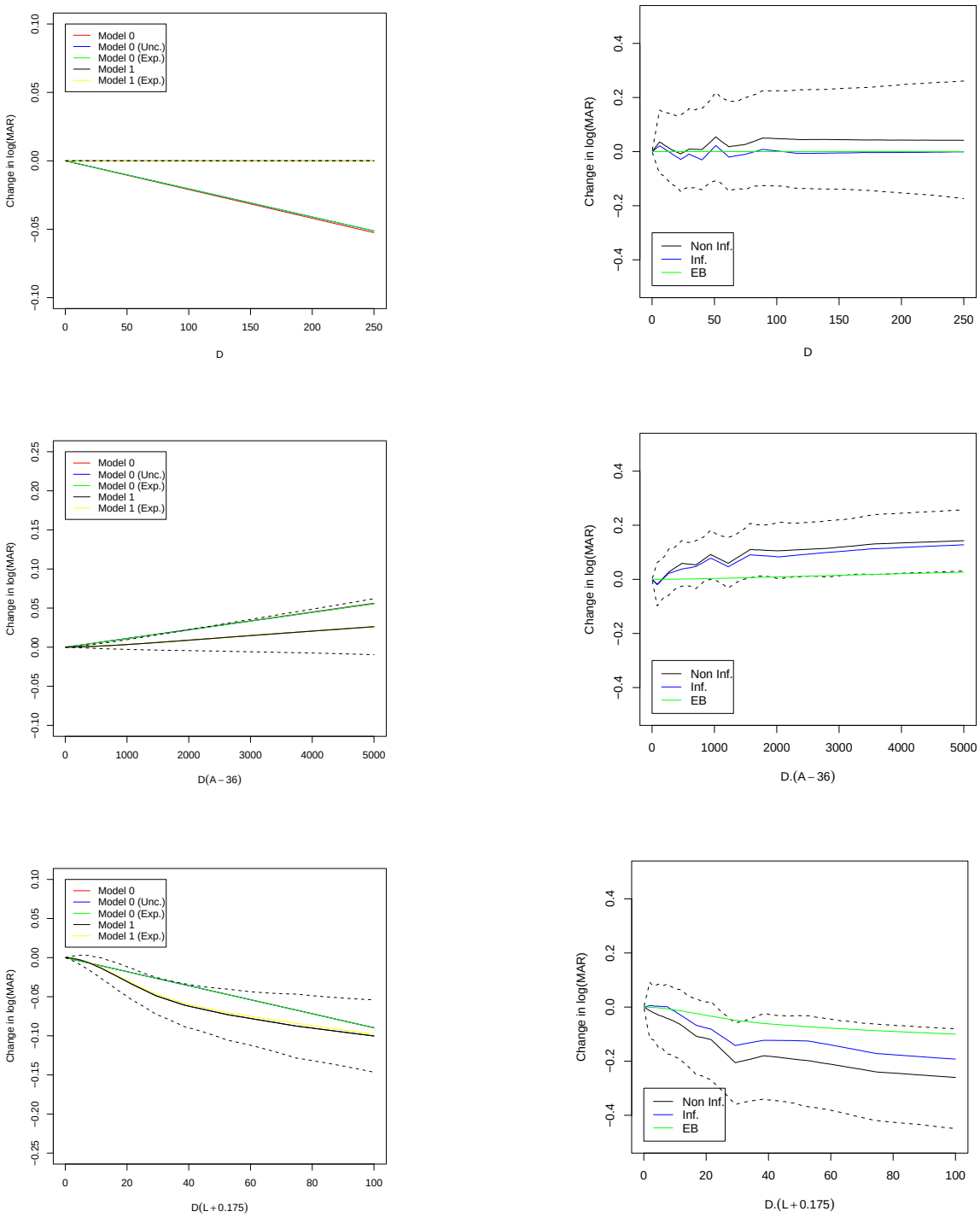

Figure 3: Estimated semiparametric functions from the linear model and SPALM analyses of the interval-level data from section 3.2. Top panel for cumulative dose D, middle panel for the interaction between D and $A - 36$, bottom panel for the interaction between D and $L + 0.175$. Left column for maximum likelihood (empirical Bayes) estimates, right column for fully Bayesian Model using Non-Informative (black) and Informative (blue) prior specification for the random effects coefficients. Empirical Bayes estimates (green) included for comparison. Dotted lines are 95% pointwise credible intervals from the Non-Informative prior model fit. Some lines overlaid.

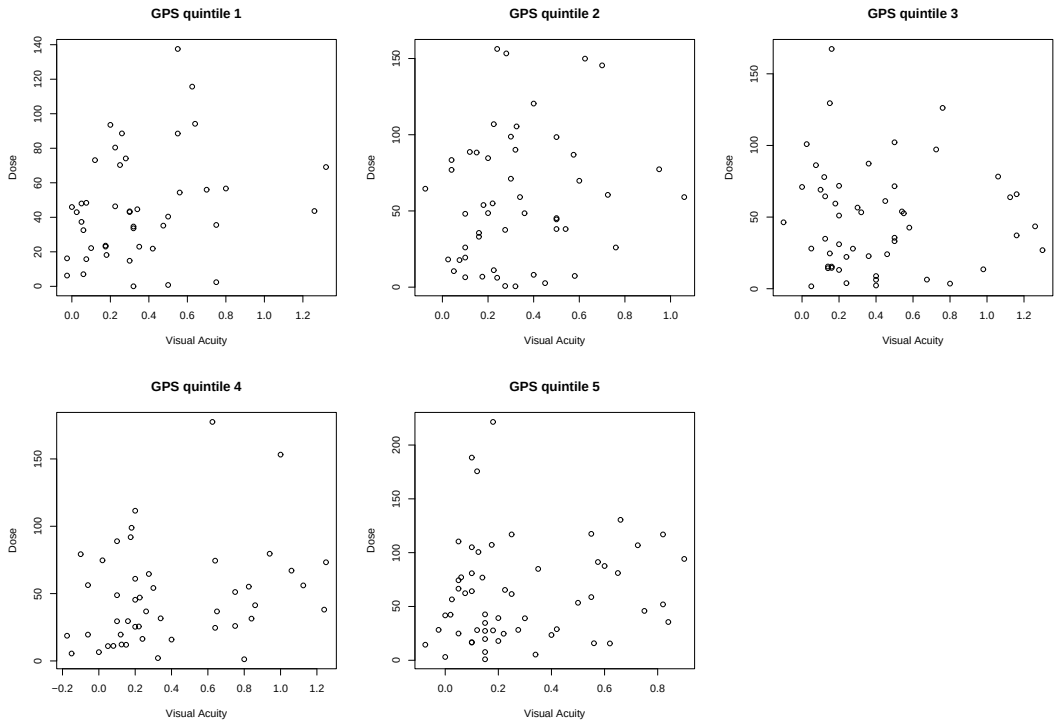

Figure 4: The balancing effect of GPS score on dose amongst observations of Visual Acuity at start of interval (see section 4.3)

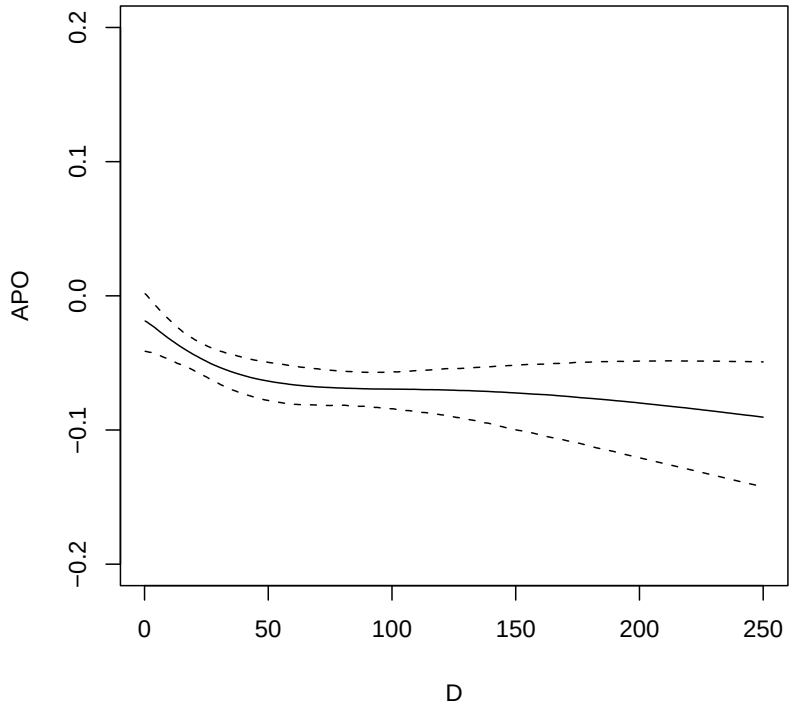

Figure 5: The estimated average potential effect for doses in the range 1 to 250 hours per interval (with approximate pointwise 95% confidence interval computed using 5000 bootstrap resamples) in the causal analysis from section 4.

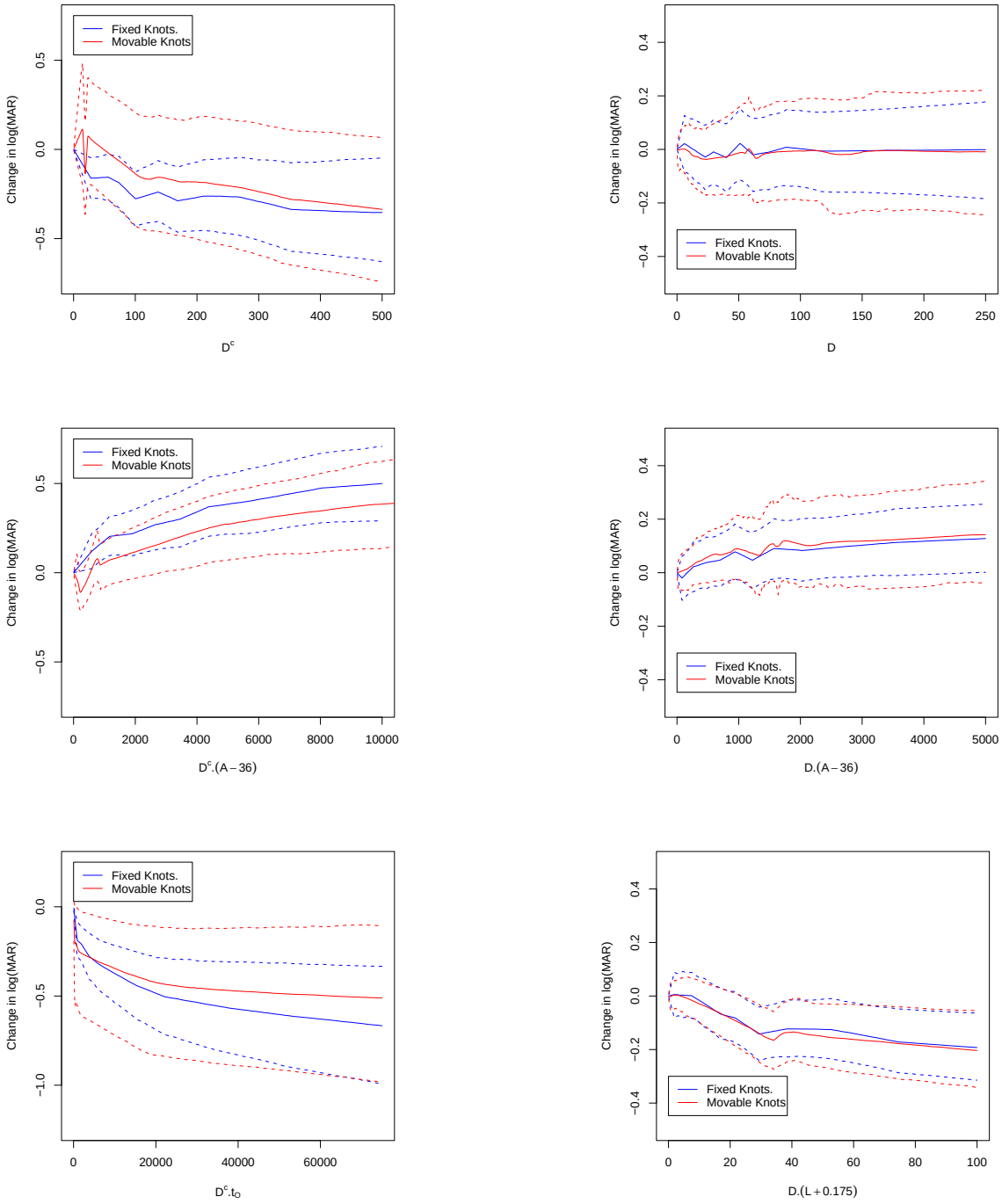

Figure 6: Estimated semiparametric functions from the informative Bayesian analyses with fixed (blue) and movable (red) knots. Left column for absolute-level data, right column for interval-level data. Dotted lines are 95% pointwise credible intervals.

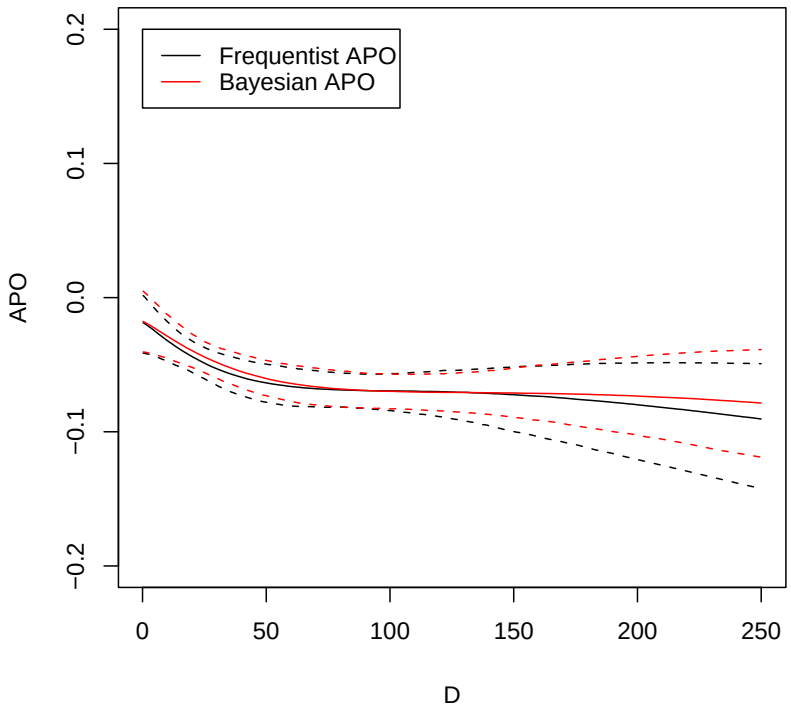

Figure 7: The estimated average potential effect for doses in the range 1 to 250 hours per interval (with approximate pointwise 95% credible interval) in the parametric Bayesian causal analysis from section 5.5. Estimated frequentist APO and 95% interval included for comparison

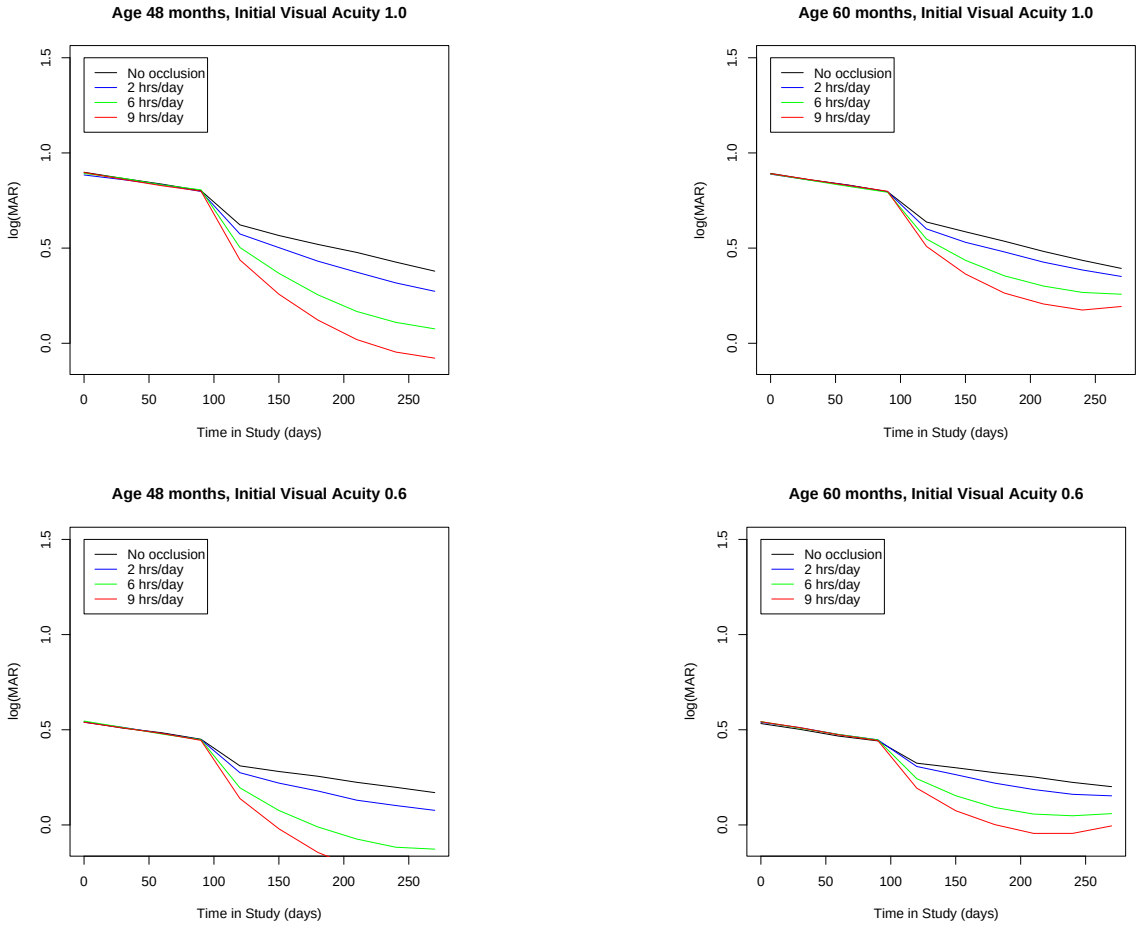

Figure 8: Predicted response for different occlusion dosing strategies in the model for longitudinal response. Profiles for anisometropic children aged 48 (left column) and 60 months (right column) at the start of study, with initial visual acuity 1.0 (top row) and 0.6 (bottom row).