Making predictions from complex longitudinal data, with application to planning monitoring intervals in a national screening programme

M. J. Sweeting* S. G. Thompson

MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge, CB2 0SR, UK.

* Corresponding author. Email: michael.sweeting@mrc-bsu.cam.ac.uk. Tel: 01223 768257. Fax: 01223 330388

Keywords: abdominal aortic aneurysm, hierarchical model, monitoring intervals, national screening, prediction, simulation

Abstract

When biological or physiological variables change over time, we are often interested in making predictions either of future measurements or of the time taken to reach some threshold value. Based on longitudinal data for multiple individuals, we develop both classical and Bayesian hierarchical models for making these predictions together with their associated uncertainty. Particular aspects addressed, which include some novel components, concern handling curvature in individuals' trends over time, making predictions for both underlying and measured levels, making predictions from a single baseline measurement, making predictions from a series of measurements, allowing flexibility in the error and random effects distributions, and including covariates. In the context of data on the expansion of abdominal aortic aneurysms over time, where reaching a certain threshold leads to referral for surgery, we discuss the practical application of these models to the planning of monitoring intervals in a

national screening programme. Prediction of the time to reach a threshold was too imprecise to be practically useful, and we focus instead on limiting the probability of exceeding the threshold after given time intervals. While more complex models can be shown to fit the data better, we find that relatively simple models seem to be adequate for the purpose of planning monitoring intervals.

1 Introduction

Interest often lies in constructing models not only for estimation of the characteristics of a longitudinal process, but also for prediction of how the process will evolve in the future. The focus of such predictions can take a number of forms: a measurement of the process at a given time in the future; the time taken to reach a certain threshold; or a probability statement about exceeding a given level at a future time. In this paper these issues are discussed in the particular context of modelling the growth of abdominal aortic aneurysms (AAAs).

AAAs are swellings in the main artery from the heart, defined as an aortic diameter of 30mm or more, which can be detected and measured by ultrasound scanning. Aneurysms that grow too large are at a substantial risk of rupture, which carries with it a high fatality rate [1]; AAAs are responsible for around 2% of all deaths in men aged over 65 [2]. It is now established practice to offer surgery if an aneurysm becomes too large and before rupture occurs, typically when the aneurysm diameter exceeds 55mm. National screening programmes have recently been established, in the UK and elsewhere [3], where men aged 65 are invited for ultrasound screening. Within such programmes monitoring intervals need to be determined so as to limit the probability of exceeding the 55mm threshold before the next scan. Clearly the length of the monitoring interval should depend on an individual's current AAA diameter, but may also be tailored to depend on other patient characteristics associated with growth rates.

The Multicentre Aneurysm Screening Study (MASS) [4] has recorded AAA diameters, using sequential ultrasound measurements, from men aged 65-74 years for up to 11 years. In this paper data from MASS are used to predict relevant quantities to help inform monitoring intervals for

the UK national screening programme. The MASS study is described in more detail in Section 4. However, to motivate the type of longitudinal model to use, Figure 1 shows the observed AAA growth for six individuals from the MASS study. These individuals were chosen to illustrate the considerable variability in growth patterns between patients and the variability within growth series. Some of the variation may be explained by patient characteristics (for example age at screening, diabetes and smoking habits [5]) although much will remain unexplained.

A flexible model is therefore required to allow for patient specific AAA growth that may be increasing or decreasing, linear or non-linear. Linear and quadratic hierarchical growth models can provide this flexibility and have been implemented previously to characterise AAA growth [5, 6, 7]. Hierarchical models (multilevel models) are commonly used to account for correlation in repeated measurements when data are hierarchically structured [8]. Such models will be used to make predictions about the future aneurysm size for a hypothetical set of individuals with one or more AAA measurements.

In Section 2 the linear and quadratic hierarchical growth models are introduced. In Section 3, given data for a specific individual, a variety of predictions are obtained from the model using the estimating random effects for that individual. The MASS dataset is described in more detail in Section 4, and predictions are obtained from hierarchical growth models fitted to the data. Finally we investigate extending the linear and quadratic models by relaxing the assumption of normally distributed random effects and allowing for a heavier tailed error distribution.

2 Linear and quadratic growth hierarchical models

Suppose repeated measurements of a variable are collected from n individuals where y_{ij} denotes the jth measurement from the ith individual, $i = 1, \ldots, n$, $j = 1, \ldots, m_i$. Measurement y_{ij} is obtained at time t_{ij}, where the time origin $t = 0$ is well-defined by, for example, a given calendar time, age, or clinical measurement. Assuming a normally distributed response, and normally

distributed intercepts and slopes, the basic form of the linear mixed effects growth model is

$$\begin{aligned} y_{ij} &= \boldsymbol{x}_{ij}(\boldsymbol{\beta} + \boldsymbol{b}_i) + \epsilon_{ij} \\ \boldsymbol{b}_i &\sim N_2(\mathbf{0}, \Sigma), \end{aligned} \tag{2.1}$$

where $\boldsymbol{x}_{ij} = (1, t_{ij})$ is the design vector for patient i at the jth measurement, $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ is a vector of fixed parameters and $\boldsymbol{b}_i = (b_{0i}, b_{1i})^T$ specify the individual specific random effects terms. The parameters β_0 and β_1 represent the average intercept and slope (rate of growth), respectively. The error terms ϵ_{ij} are assumed independent $N(0, \sigma_w^2)$, whilst the between-subject variance-covariance matrix Σ has variances σ_0^2 and σ_1^2 on the diagonal, and covariance $\rho_{01}\sigma_0\sigma_1$ on the off-diagonal. In addition, suppose p covariates are available from individual i at measurement j, given by the p-dimensional vector $\boldsymbol{z}_{ij}$. If the effect of these covariates on both the intercept and slope is of interest, then the covariate design vector $\boldsymbol{w}_{ij} = (\boldsymbol{z}_{ij}, t_{ij}\boldsymbol{z}_{ij})$ can be formed to give the following model:

$$y_{ij} = \boldsymbol{x}_{ij}(\boldsymbol{\beta} + \boldsymbol{b}_i) + \boldsymbol{w}_{ij}\boldsymbol{\gamma} + \epsilon_{ij}, \tag{2.2}$$

where $\boldsymbol{\gamma}$ is a $2p \times 1$ vector of fixed parameters containing the effect of each covariate on the intercept and slope.

The model can be extended to allow for curvature in individual growth rates by considering a quadratic growth model. The design vector then becomes $\boldsymbol{x}_{ij} = (1, t_{ij}, t_{ij}^2)$ with corresponding fixed effect parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$, and random effects $\boldsymbol{b}_i = (b_{0i}, b_{1i}, b_{2i})^T$ modelled using a trivariate Normal distribution. Σ is now a 3×3 variance-covariance matrix, with variances σ_0^2, σ_1^2, and σ_2^2, and covariances $\rho_{kl}\sigma_k\sigma_l$, $k, l = 0, 1, 2$, $k \neq l$. The effect of the p covariates on the intercept, slope, and curvature can be estimated using the covariate design vector $\boldsymbol{w}_{ij} = (\boldsymbol{z}_{ij}, t_{ij}\boldsymbol{z}_{ij}, t_{ij}^2\boldsymbol{z}_{ij})$ with associated $3p \times 1$ vector $\boldsymbol{\gamma}$. We shall let $\theta = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \Sigma, \sigma_w)$ denote the vector of parameters that are to be estimated by the model.

Growth models in which the random effects are normally distributed can be fit using maximum likelihood or restricted maximum likelihood (REML) in a number of statistical packages (`xtmixed` (Stata), `gllamm` (Stata) `nlme/lme4` (R), `proc mixed` (SAS)). Alternatively, a Bayesian approach

can be taken where the parameters θ are given prior distributions, and posterior inferences obtained via Markov chain Monte Carlo (MCMC) [9].

3 Hierarchical model predictions

Predictions of AAA growth can be used to inform the length of monitoring intervals for AAA screening. Such predictions may be made from a fitted hierarchical model for a number of possible hypothetical individuals, not in the original dataset (an out-of-sample prediction). For example, it may be of interest to predict the diameter of an aneurysm, say one year after a screening measurement of 40mm is taken. Alternatively, a patient may have two or more repeat ultrasounds recorded, and all such measurements may then be used to make future predictions. One important question is whether the current diameter is adequate to make a precise prediction or whether repeated measurements are required.

This section deals with the situation where predictions are to be made for a specific individual given one or more response measurements with corresponding times of measurement. The predictions use random effects which have been estimated conditional on the individual's likelihood. Suppose $\boldsymbol{b}_i$ are the random effects estimated from a linear model for individual i with data $(\boldsymbol{y}_i, \boldsymbol{t}_i)$. Within a classical framework, the conditional modes of the random effects, also referred to as best linear unbiased predictors (BLUPs), can be obtained for an individual from most software packages. However, the distribution of the random effects given the observed data will generally be of a non-standard form. If however Bayesian MCMC is used, it is easy to obtain the posterior distribution of the individual specific random effects. Furthermore, if the `cut` function in WinBUGS is used [11], the random effects for a new individual can be estimated without the likelihood from this individual's data updating the population parameters of the model. In what follows every prediction is a function of the fixed effects, random effects and possibly the measurement error.

3.1 The estimated rate of growth

For a linear model the rate of growth for individual i, $G_i^L(\theta) = \beta_1 + b_{1i}$, is constant over time, whilst for the quadratic model the rate of growth at time t is $G_i^Q(t;\theta) = (\beta_1 + b_{1i}) + 2t(\beta_2 + b_{2i})$. The posterior distribution for G_i can be easily calculated using Bayesian MCMC, and inferences made from this distribution.

3.2 Prediction of time taken to cross a threshold given a current measurement

One prediction that may be of interest is the time taken for an individual's underlying growth curve to cross a certain threshold, α, from any given time t. For the linear growth model without covariates the time taken for individual i to hit threshold α can be calculated as:

$$W_i^L(t,\alpha;\theta) \quad = \quad \frac{\alpha - \beta_0 - b_{0i}}{\beta_1 + b_{1i}} - t. \tag{3.1}$$

One complication with the variable W_i^L is that it may take negative values. This could happen either because the individual is already over the threshold at time t, or because the true growth rate is negative and hence the threshold was crossed in the past. The first case occurs when $(\beta_0+b_{0i})+(\beta_1+b_{1i})t \geq \alpha$ and the second when $\beta_1+b_{1i} < 0$. If, as is usually the case, the primary question concerns the time till the process is greater than or equal to α, then it is necessary to evaluate W_i^L as zero if $(\beta_0+b_{0i})+(\beta_1+b_{1i})t \geq \alpha$, since the process is already above the threshold. On the other hand, if $(\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t < \alpha$ and $\beta_1 + b_{1i} < 0$ then W_i^L will be infinite, since the threshold will never be crossed in the future.

Using a quadratic model, predictions of this type are even more complex. For a given time t, interest lies in the *first* time in the future at which the threshold is crossed. As with the linear model we should first assess whether $(\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t + (\beta_2 + b_{2i})t^2 \geq \alpha$, and if so set $W_i^Q = 0$. Otherwise, W_i^Q should be calculated as:

$$W_i^Q(t,\alpha;\theta) = T_i(\alpha;\theta) - t, \tag{3.2}$$

where $T_i(\alpha;\theta)$ is the first time *after* t at which the threshold is crossed. This can be calculated from the following equation:

$$T_i(\alpha;\theta) = \frac{-(\beta_1 + b_{1i}) \pm \sqrt{(\beta_1 + b_{1i})^2 - 4(\beta_2 + b_{2i})(\beta_0 + b_{0i} - \alpha)}}{2(\beta_2 + b_{2i})}. \tag{3.3}$$

If there are no roots to this equation, then the quadratic curve will never cross the threshold, and hence $W_i^Q(t,\alpha;\theta) = \infty$ for all values of t. Likewise, if both roots to this equation are less than t then the threshold will not be crossed again, and $W_i^Q(t,\alpha;\theta) = \infty$. Otherwise, we take the first root of $T_i(\alpha;\theta)$ that occurs after t. Using these rules, posterior distributions for W_i^L and W_i^Q can be obtained using Bayesian MCMC. In a classical analysis, the properties of this random variable are far harder to compute, and some simulation technique would be required.

3.3 Prediction of a measurement at a given future time

Predictions of future measurements can be obtained relatively easily. For the linear growth model without covariates, the predicted measurement at time $t + s$ is simply

$$Y(t,s;\theta) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})(t+s) + \epsilon, \tag{3.4}$$

whilst for the quadratic model the estimated measurement at time $t + s$ is

$$Y(t,s;\theta) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})(t+s) + (\beta_2 + b_{2i})(t+s)^2 + \epsilon. \tag{3.5}$$

The posterior distribution of this predictive quantity is more commonly known as the posterior predictive distribution. Importantly, to obtain the probability that a measurement is above a certain value at a given future time we can simply calculate the tail area of the posterior predictive distribution corresponding to that which is above the chosen value.

4 The Multicentre Aneurysm Screening Study (MASS)

MASS was set up to assess whether or not screening for AAA was beneficial in terms of long-term mortality [4]. Between 1997 and 1999, men aged 65-74 years were recruited from family doctor

lists in four UK centres. Of the 33883 men invited to screening, 26875 had a visualised abdominal ultrasound scan and 1334 aneurysms (diameter $\geq$ 30mm) were detected. For this analysis of growth rates, data are taken from 1046 subjects who had a diameter 30-54mm at their first screen and at least one follow-up ultrasound measurement. The current aneurysm diameter determined the next examination time; individuals who measured 30-44mm were rescanned a year later, whilst those with diameters 45-54mm were rescanned after a further 3 months. In total, the data contains 8941 ultrasound examinations. The average duration of follow-up was 4.9 years, with a mean of 8.5 ultrasound scans per person.

4.1 Follow-up and censoring

Individual series are terminated either due to surgery (36%), death (21%), loss to follow-up (26%), or the administrative censoring date of 31st March 2008 (17%), whichever comes first. Individuals whose aneurysm diameter measured 55mm or greater at any examination or who showed rapid expansion (defined as observed growth $\geq$ 10mm in one year), were considered for elective surgery. Those deemed unsuitable for surgery had continued surveillance of their aneurysm. A series that is terminated due to the patient undergoing elective surgery will tend to be biased towards a larger diameter on the final measurement due to measurement error[5]. However, the drop-out process is ignorable if the analysis uses a likelihood-based hierarchical model since drop-out depends on the observed data (missing at random) [12]. Figure 2 shows four spaghetti plots of individual growth series, grouped by the mode of termination, together with the empirical mean AAA diameter profiles. In this figure only measurements taken close to an anniversary of screening are used, since 3 and 6 month re-scans were only undertaken in individuals with diameters 45-54mm, and could skew results. It can clearly be seen that on average AAA diameters are larger in the group who eventually go for surgery, and those who become lost to follow-up have on average smaller AAAs. Due to this diameter dependent drop-out the empirical mean AAA diameter profile should always be interpreted with caution [12]. Nevertheless, there is no reason to suspect that drop-out is based on information not observed in the measurement process, and hence for this analysis we

assume (missing at) random drop-out.

Of the 3846 non-final ultrasounds that measured 30-44mm, 3636 (95%) had a repeat measurement within 14 months, broadly following protocol. 4041 non-final ultrasounds measured 45-54mm, for which 2994 (74%) had a repeat measurement within 5 months. The more regular 3-month follow-up appointments were therefore less strictly adhered to, either due to the patient not attending or due to the appointment not being scheduled. The effect of these missed appointments on the analysis should be minor, since these data are only intermittently missing and the missingness is believed to be random.

4.2 Estimation

Both classical REML and Bayesian MCMC are used to obtain estimates of the parameters. Non-informative priors are used for the Bayesian models. The population mean parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ are given vague independent $N(0, \tau^2)$ priors with $\tau = 1000$. The within-subject variance σ_w^2 is assigned an inverse-gamma prior, $IG(0.001, 0.001)$. To ensure that Σ is positive-definite, an inverse-Wishart prior distribution is used with degrees of freedom equal to one plus the the dimension of Σ, *i.e.* 3 for the linear model, and 4 for the quadratic. This has the effect of placing a uniform distribution on each of the correlation parameters [9].

4.3 Time scale for analysis

There are two possible choices for the time scale used in the longitudinal model; time since screening and age. Time since screening is relevant since at baseline the population are constrained to be within the diameters 30-54mm; the inclusion policy of the MASS study. This is also the inclusion criteria for the UK National Screening Programme [TRUE??], and hence this time scale is highly relevant for predictions. However, using age as the time scale may be more relevant for general predictions of aneurysm growth, where the time of screening is an irrelevant quantity. A comparison of models using each time scale was first made. In a hierarchical model, the choice of time scale is important as shrinkage of the random effects can result in different estimates of

mean growth, and can change predictions. This is seen in Table 1, where estimates from classical linear and quadratic growth models, using either time since screening or age as the time scale, are presented (linear models: L1-time and L1-age; quadratic models: Q1-time and Q1-age). The fixed-effect estimates of mean AAA growth are quite different between the models L1-time and L1-age, and between Q1-time and Q1-age.

The models can be further compared by studying the AIC. Clearly a nonlinear trend provides a better fit as the AIC decreases dramatically in the two quadratic models. Furthermore, the use of time since screening as the time scale provides a better fit to the MASS data. In terms of prediction, AIC helps us to choose a model that will give good predictions for a new individual recruited in the same way as the sample, and clearly the models that use time since starting are better in this respect. Time since screening is therefore used as the time scale in all following models, but to make relevant predictions for the national screening programme we also consider including baseline age as a covariate in Section 4.5. This facilitates predictions to be made for a number of possible ages at screening, and in particular age 65.

4.4 Bayesian models

Table 2 shows the parameter estimates for the standard linear (L1) and quadratic (Q1) models, fit using Bayesian MCMC fit. Compared with the maximum likelihood estimates obtained from the classical fit (Table 1), both the classical and Bayesian models produce almost identical parameter estimates suggesting that the priors chosen in the Bayesian models are indeed effectively non-informative. Table 2 also shows the posterior mean deviance $\bar{D}$, the effective number of parameters p_D and the deviance information criterion, DIC [13]. From model L1, the average diameter at first screen is 37.5mm (SE=0.2), with an average growth rate of 2.2mm/year (SE=0.07). There is considerable between-patient variation both in AAA diameters at first screen and in growth rates, and these are positively correlated. As with the classical models, there is evidence that AAA growth is non-linear since the quadratic model (Q1) has a lower DIC.

Figure 3 shows the distribution of measured aneurysm diameters at first screen. Clearly

the distribution is skewed and non-normal, indicating that the standard model may be inadequate. One alternative is to allow the baseline intercepts to be lognormal, $\beta_{0i} \equiv \beta_0 + b_{0i} \sim$ Lognormal$(\mu_{LN}, \sigma^2_{LN})$, so that $E[\beta_{0i}] = \beta_0 = \exp(\mu_{LN} + \sigma^2_{LN}/2)$ and $V(\beta_{0i}) = \sigma^2_0 = (\exp(\sigma^2_{LN}) - 1)\exp(2\mu_{LN} + \sigma^2_{LN})$. The random slopes can then be modelled conditionally on the random intercepts by assuming the conditional distribution is Gaussian, as follows:

$$\begin{aligned} \beta_{1i}|\beta_{0i} &\sim N(\mu_{Ci}, \sigma^2_C) \\ \mu_{Ci} &= \beta_1 + \lambda(\beta_{0i} - \beta_0) \\ \sigma^2_C &= \sigma^2_1 - \lambda^2\sigma^2_0. \end{aligned} \tag{4.1}$$

This parameterisation results in $E[\beta_{1i}] = \beta_1$ and $V(\beta_{1i}) = \sigma^2_1$. This model, L2, has a somewhat superior DIC to L1 (Table 2).

To avoid making a parametric assumption concerning the distribution of diameters at first screen, model L3 allows the individual specific intercepts to be 'free' fixed effects which have no hierarchical structure imposed. Hence this model estimates 1046 separate intercepts with no shrinkage towards their overall mean. The slopes are estimated as conditional on the intercepts (as described in Equation (4.1)). Since the population of the intercepts is not specified and hence β_0 and σ_0 are not parameters of the model, in Table 2 the unweighted empirical means and variances of the intercepts are presented. The standard deviation, σ_0 is higher than estimated in models L1 and L2 reflecting the fact that no shrinkage of the intercepts is taking place. Conversely the standard deviation of the slopes σ_1 is smaller as is the empirical correlation between intercepts and slopes. Interestingly, the effective number of parameters only increases by 53 compared to the random effects model L1 and the posterior mean deviance, $\bar{D}$, is actually higher in model L3, as is the DIC. One possible reason for the increase in $\bar{D}$ is that, since the deviance is averaged over its posterior distribution, $\bar{D}$ already incorporates a degree of penalty for model complexity. Indeed the relatively small increase in the effective number of parameters (compared to the 1046 individuals) suggests that there is not much shrinkage of the intercepts under model L1. Nevertheless, the smaller DIC in L1 indicates that this is the preferred model.

There is evidence from residual plots that the within-patient variation is more heavy tailed than Gaussian. So a further model that we consider specifies a t-distribution for within-patient variation. The degrees of freedom of this distribution are to be estimated, and we place a Uniform(2,1000) prior on the degrees of freedom parameter. Results from this extension to the linear model, labelled L1-T, are given in Table 2. The degrees of freedom are estimated to be close to four, suggesting a heavy tailed distribution, and the DIC has decreased substantially compared with model L1.

Table 3 shows predictions for a specific individual whose AAA diameter at screening ($t = 0$) is either 35mm or 50mm. All models estimate a similar true growth rate when $y = 35$ at first screen, at approximately 2mm/year. The predicted growth rates when $y = 50$ at first screen are however higher, at approximately 3.5mm/year. The estimated time for the underlying process to cross 55mm is similar across all models, although the wide credible intervals limit practical usefulness of this quantity. Figure 4 shows predicted aneurysm growth given a single measurement at screening of either (a) 35mm or (b) 50mm. In general, predictions are remarkably similar between the fitted models, with the quadratic model, Q1, showing slight curvature for an individual with a 35mm diameter at screening. Interestingly, for an individual who measures 50mm at screening, the predicted average AAA diameter 3 months later is slightly less than 50mm for all models except L3. This occurs because the intercepts from all these models are shrunk towards the population mean intercept, resulting in slightly lower predictions, whilst there is no shrinkage of intercepts in model L3. For an individual with a 30mm diameter at screening, predictions are more similar since the observed diameter is closer to the population mean diameter resulting in less shrinkage.

In terms of planning monitoring intervals for AAA, a key desire is to limit the probability that the next observation is over the 55mm threshold. Such probabilities can be easily calculated from the predictive distributions in an MCMC framework, and Figure 5 shows how these depend on the baseline AAA diameter and can be controlled by choosing the time of next measurement. Both the linear and quadratic models are shown in the figure for probability limits of 1%, 5%, and 10%. For example, if we wish for fewer than 1% of individuals to have a diameter over the threshold

at their second scan, a screening interval of 2.5 years or less would be sufficient for those who measured 35mm at baseline. In contrast this interval would need to be six months or less for an individual who measured 45mm at baseline. For individuals who measured 50mm at baseline there is actually already a greater than 1% chance that an immediate re-measurement would result in an observed diameter over 55mm. The linear and quadratic models give very similar results apart from when the AAA diameter is small and the probability that we wish to limit is relatively large (*e.g.* 10%).

4.5 Predictors of AAA growth

We consider extending model L1 to include possible predictors of AAA growth. At first screen individuals were asked about their current smoking habits. Ninety-seven individuals reported never smoking compared to 585 previous smokers and 317 current smokers. Smoking data was missing for 47 individuals. The population parameters for this model are very similar to model L1 although there is strong evidence that previous and current smokers have, on average, larger diameters at baseline than non-smokers (by 2.4mm (SE=0.8) and 2.5mm (SE=0.9), respectively), and faster growth than non-smokers (by 0.53mm/year (SE=0.21) and 0.82mm/year (SE=0.22), respectively). The age of an individual at baseline was also considered as a predictor of aneurysm growth. There was found to be no evidence of an association between age and AAA size at screening (-0.08mm, SE=0.08), and only a mild effect between age and the rate of AAA growth (-0.04mm/year, SE=0.02). The negative coefficient suggests smaller AAA growth in the older population. However, this surprising effect may be due to the MASS selection process. One hypothesis is that fast growers in the older population will have diameters too large to be included in MASS, whilst slow growers in the young population have diameters too small for selection. Such a selection bias could produce an apparently negative association between age and growth.

Figure 6 shows how predictions vary when the smoking status of an individual is accounted for, and when predictions are based on either two, three or five measurements taken over the first year after screening. All predictions are shown for an individual aged 65 at screening. Predictions

are made for three 'slow' growers who have AAA measurements of 45mm at $t = 0$ and $t = 1$, and for three 'fast' growers who have AAA measurements of 42.5mm at $t = 0$ and 47.5mm at $t = 1$. The mean observed AAA diameter for each individual is the same (45mm). Predictions change only very slightly between the smoking categories, despite the highly significant effect of including this variable as a covariate in the model. Furthermore, predictions based on a 'fast' or 'slow' grower are also quite similar for years 2 and 3 after screening, although they start to diverge after years 4 and 5. Surprisingly, the number of measurements does not significantly alter the precision of the estimates, which are almost indistinguishable between patients who have two measurements compared with those who have five. For 'slow' growers, the predicted growth rate is higher than that observed, whilst for 'fast' growers the predicted growth rate is less than that observed. Finally, all the predicted growth curves appear to pass close to the average observed diameter (45mm) at the average observation time ($t = 0.5$).

5 Discussion

We have shown in this paper how various predictions can be made using a linear or quadratic hierarchical mixed effects model. Two novel aspects arise from this work. Firstly, we have extended the mixed effects models to incorporate error and random effects distributions that are non-normal, within a Bayesian framework. Secondly, we make predictions for a specific individual using random effects estimated conditional on specific likelihood contributions.

A slightly different prediction approach has been described by Skrondal and Rabe-Hesketh [14]. Here, the population-averaged predicted mean response is obtained analytically by integrating over the random effects distribution, whilst uncertainty in the mean response is addressed by simulating the parameters from their sampling distribution. A parametric bootstrap procedure has also been proposed as a way of obtaining a prediction interval for the mean response given values of the covariates [15]. Meanwhile, Taylor and Law give an account of how in-sample predictions of future observations can be obtained from closed form solutions when multivariate Normality is assumed

[16].

One practical advantage of using Bayesian modelling within a flexible software package such as WinBUGS is that many different model extensions can easily be investigated. For example, it is clear that for an aneurysm detected population the baseline distribution of aneurysm diameters is non-normal. We have tried to relax the normality assumption placed on our random effects by fitting a model with lognormal intercepts. However even this model is not entirely adequate. One alternative approach is to model more precisely the process by which the data were obtained. Specifically, individuals who were screened and deemed 'normal', that is had a diameter <30mm, were not followed-up further and hence were excluded from the analysis dataset. Therefore in reality, in addition to the 1046 measured AAA diameters at first screen, there are also 25541 left censored diameters, in which we only know that $y < 30$. Each of these individuals have the following likelihood contribution:

$$L_i = \int p(y_{i1} < 30|\theta, b_{0i})p(b_{0i}|\theta)\, db_{0i}.$$

By adding these contributions to the likelihood, our inferences are then about the general population, and not specifically about those with an aneurysm. As a consequence, for all individuals with a detected aneurysm, their estimated baseline diameters are likely to be shrunk downwards towards the population mean. For example, an individual whose observed diameter is 30mm at first screen is more likely to have their true diameter less than 30mm, due to our knowledge that the population mean aneurysm diameter is far smaller. This behaviour can only be modelled if the censoring mechanism is fully incorporated. We have attempted to fit censoring models by assuming that the population distribution of AAA diameters at first screen follows either a Gaussian distribution or t-distribution with the degrees of freedom estimated by the model. The population mean diameter at first screen was estimated to be close to 20mm whilst the population mean rate of growth was 0.05mm/year for the model with Gaussian intercepts, and 1.18mm/year for the model with t-distributed intercepts. Hence, estimates from these models are highly sensitive to the choice of distribution for the intercepts. This is due to the fact that 95% of the individuals are censored and only measurements for the upper 5% tail of the distribution are available.

To our knowledge, the issue of such censoring/truncation has been rarely addressed when using longitudinal mixed effects models. Mehrotra and colleagues proposed an EM-like algorithm to obtain maximum likelihood estimates when the sample is truncated, but only for a fixed subject effects model [17]. Further investigation into the behaviour of mixed effects models when censoring/truncation is present would therefore be of interest.

The non-linear growth of AAAs has been shown previously [5, 7]. The use of mixed effects models with correlated intercept and linear growth rates allows individuals with higher baseline measurements to have faster growth. Accelerated growth within an individual's growth series can also be modelled using a non-linear model. However, we have shown that using either a quadratic or linear model gave remarkably similar predictions over the time period of interest for AAA monitoring. Since there are difficulties associated with making predictions from a quadratic model, we would question the practical relevance of this model for this application. Indeed quadratic growth may be unrealistic in the long-term, with predictions possibly showing a reversal in the direction of growth for some individuals. The linear model, despite representing a simplified version of the true nature of AAA growth, appears to be adequate for short-term predictions.

In all the models fitted there was found to be very substantial between-individual variation, which requires further exploration. The baseline smoking status of an individual was found to be significantly associated with both baseline AAA diameter and the rate of growth. Nevertheless, the between-individual standard deviations for the intercept and slope only decreased by 0.4% as a result of adding smoking status as a covariate. Other variables that have been shown to correlate with AAA growth rates include diabetes and atherosclerosis [5]; although such variables could be included in future models, it is unlikely they would impact importantly on relevant predictions.

A variety of predictions can be made from longitudinal models, such as the time to reaching a certain threshold, or the predicted level of the observed or underlying outcome after a given time period. In our AAA application, however, we find that a prediction of the time taken to reach a threshold diameter of 55mm is of little practical use, since the prediction is very imprecise. This has been noted previously in relation to time-to-event predictions in the context of survival analysis

[18]. Joint longitudinal data and survival modelling [19] are inappropriate in our application, since we are modelling the time to an underlying threshold that is an aspect of the longitudinal process, rather than to an observed event. More relevant for planning monitoring intervals is the distribution of the (observed) outcome after a given time, and the probability that a future observation will be greater than a specified threshold at that time. Expressing the prediction in terms of the probability of crossing the threshold provides a rational basis for planning appropriate monitoring intervals.

Acknowledgements

We would like to thank Hilary Ashton and Lu Gao for providing the MASS data and resolving queries, and David Spiegelhalter and Vern Farewell for providing useful comments regarding the manuscript and truncated baseline diameters. ST and MS are employed by the UK Medical Research Council (MRC grant U.1052.00.001 and NIHR HTA grant 08/30/02).

References

[1] M. J. Bown, A. J. Sutton, P. R. F. Bell, and R. D. Sayers. A meta-analysis of 50 years of ruptured abdominal aortic aneurysm repair. *British Journal of Surgery*, 89:714–730, 2002.

[2] Office of National Statistics. Mortality Statistics Series DH2 no. 27. Deaths in England and Wales, 2000.

[3] UK National Screening Committee. Abdominal aortic aneurysm screening, May 2007. Accessed at www.library.nhs.uk/screening/.

[4] S. G. Thompson, H. A. Ashton, L. Gao, and R. A. P. Scott. Screening men for abdominal aortic aneurysm: 10 year mortality and cost effectiveness results from the randomised Multicentre Aneurysm Screening Study. *British Medical Journal*, 338:b2307, 2009.

[5] A. R. Brady, S. G. Thompson, F. G. R. Fowkes, R. M. Greenhalgh, J. T. Powell, and UK Small Aneurysm Trial Participants. Abdominal aortic aneurysm expansion - Risk factors and time intervals for surveillance. *Circulation*, 110:16–21, 2004.

[6] P. Eriksson, S. Jormsjo-Pettersson, A. R. Brady, H. Deguchi, A. Hamsten, and J. T. Powell. Genotype-phenotype relationships in an investigation of the role of proteases in abdominal aortic aneurysm expansion. *British Journal Of Surgery*, 92:1372–1376, 2005.

[7] K. A. Vardulaki, T. C. Prevost, N. M. Walker, N. E. Day, A. B. M. Wilmink, C. R. G. Quick, H. A. Ashton, and R. A. P. Scott. Growth rates and risk of rupture of abdominal aortic aneurysms. *British Journal Of Surgery*, 85:1674–1680, 1998.

[8] H. Goldstein. *Multilevel Statistical Models*. London: Edward Arnold, third edition, 2003.

[9] A. Gelman and J. Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, 2007.

[10] Douglas Bates and Martin Maechler. *lme4: Linear mixed-effects models using S4 classes*, 2009. R package version 0.999375-32.

[11] D. Spiegelhalter, A. Thomas, N. Best, and Lunn D. *WinBUGS Version 1.4 User Manual*. MRC Biostatistics Unit, Cambridge, 2003.

[12] P. Diggle, P. Heagerty, K. Liang, and S. Zeger. *Analysis of longitudinal data*. Oxford: Oxford University Press, second edition, 2002.

[13] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit. *Journal Of The Royal Statistical Society Series B-Statistical Methodology*, 64:583–616, 2002.

[14] A. Skrondal and S. Rabe-Hesketh. Prediction in multilevel generalized linear models. *Journal Of The Royal Statistical Society Series A-Statistics In Society*, 172:659–687, 2009.

[15] P. Hall and T. Maiti. On parametric bootstrap methods for small area prediction. *Journal Of The Royal Statistical Society Series B-Statistical Methodology*, 68:221–238, 2006.

[16] J. M. G. Taylor and N. Law. Does the covariance structure matter in longitudinal modelling for the prediction of future CD4 counts? *Statistics In Medicine*, 17:2381–2394, 1998.

[17] K. G. Mehrotra, P. M. Kulkarni, R. C. Tripathi, and J. E. Michalek. Maximum likelihood estimation for longitudinal data with truncated observations. *Statistics In Medicine*, 19:2975–2988, 2000.

[18] E. Graf, C. Schmoor, W. Sauerbrei, and M. Schumacher. Assessment and comparison of prognostic classification schemes for survival data. *Statistics In Medicine*, 18:2529–2545, 1999.

[19] A. A. Tsiatis, V. De Gruttola, and M. S. Wulfsohn. Modelling the relationship of survival to longitudinal data measured with error - application to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association*, 90:27–37, 1995.

Parameter	L1-time	L1-age*	Q1-time	Q1-age*
β_0	37.5 (0.2)	38.3 (0.3)	38.3 (0.2)	39.3 (0.3)
β_1	2.19 (0.06)	1.89 (0.05)	1.48 (0.09)	1.24 (0.06)
β_2	-	-	0.109 (0.009)	0.106 (0.006)
σ_0	7.13	9.71	6.69	8.87
σ_1	1.70	1.43	2.27	1.62
σ_2	-	-	0.16	0.09
ρ_{01}	0.52	0.34	0.58	0.67
σ_w	3.12	3.16	2.96	3.03
$-2\log(L)$	-25677	-26018	-25444	-25732
k	6	6	10	10
AIC	51366	52049	50979	51555

Table 1: Parameter estimates (SE) from classically fit linear and quadratic growth models, using either time since screening or age as the time scale. * Age centred at 70 years.

Parameter	L1	Q1	L2	L3	L1-T
β_0	37.5 (0.2)	38.3 (0.2)	37.5 (0.2)	37.5*	37.4 (0.2)
β_1	2.19 (0.07)	1.49 (0.09)	2.19 (0.06)	2.13 (0.06)	2.18 (0.06)
β_2	-	0.108 (0.012)	-	-	-
σ_0	7.12 (0.17)	6.69 (0.16)	7.00 (0.18)	7.69*	7.25 (0.17)
σ_1	1.74 (0.06)	2.26 (0.10)	1.71 (0.06)	1.54 (0.06)	1.65 (0.06)
σ_2	-	0.15 (0.01)	-	-	-
ρ_{01}	0.51 (0.03)	0.58 (0.04)	0.51 (0.03)	0.41*	0.51 (0.03)
σ_w	3.12 (0.03)	2.97 (0.03)	3.11 (0.03)	3.14 (0.03)	3.25 (0.07)
$\bar{D}$	45692	44823	45681	45810	44589
p_D	1620	1812	1603	1673	1723
DIC	47312	46635	47284	47483	46312

Table 2: Parameter estimates from Bayesian linear and quadratic random effects models. See text (Section 4.4) for description of models. Posterior medians with standard deviations in parentheses are shown. * Empirical means and variances.

Baseline diameter	Prediction	L1	Q1	L2	L3	L1-T
35mm	G (mm/year)	1.9	2.1	1.9	1.9	2.0
		(-1.0, 5.0)	(0.4, 5.1)	(-1.1, 4.9)	(-0.9, 4.8)	(-0.9, 4.8)
	$W(t=0,\alpha=55)$ (years)	10.1	9.1	10.5	10.3	10.1
		(3.5, ∞)	(3.7, ∞)	(3.5, ∞)	(3.6, ∞)	(3.6, ∞)
50mm	G (mm/year)	3.5	3.6	3.5	3.2	3.5
		(0.4, 6.5)	(1.0, 7.3)	(0.6, 6.6)	(0.4, 6.0)	(0.5, 6.3)
	$W(t=0,\alpha=55)$ (years)	2.0	2.0	2.0	1.8	1.9
		(0.4, 39.0)	(0.4, 10.5)	(0.3, 58.8)	(0.2, ∞)	(0.5, 23.2)

Table 3: Predictions for a specific individual given a single AAA diameter measurement at baseline ($t = 0$).

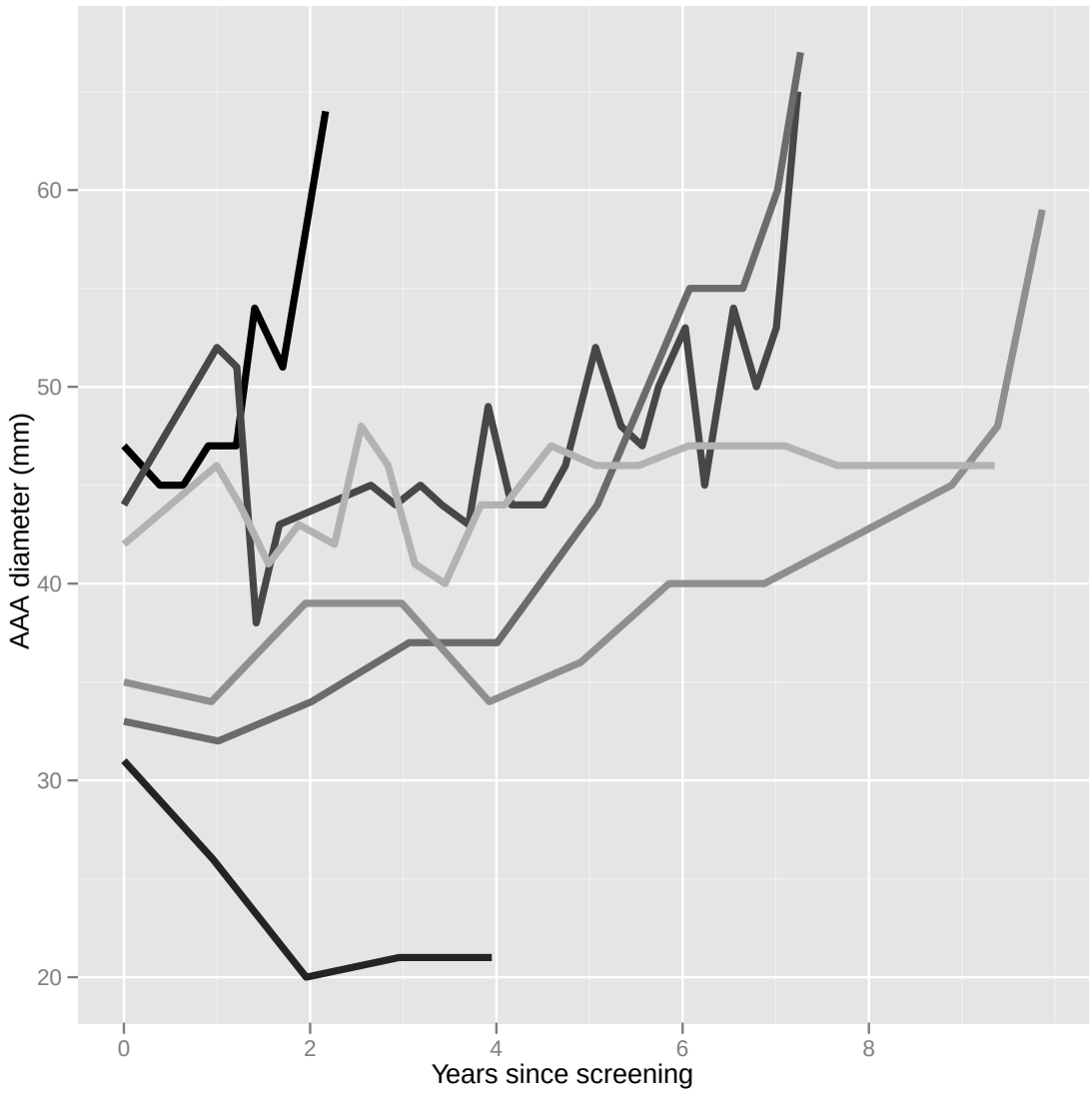

Figure 1: Abdominal aortic aneurysm growth trajectories for six individuals from the Multicentre Aneurysm Screening Study.

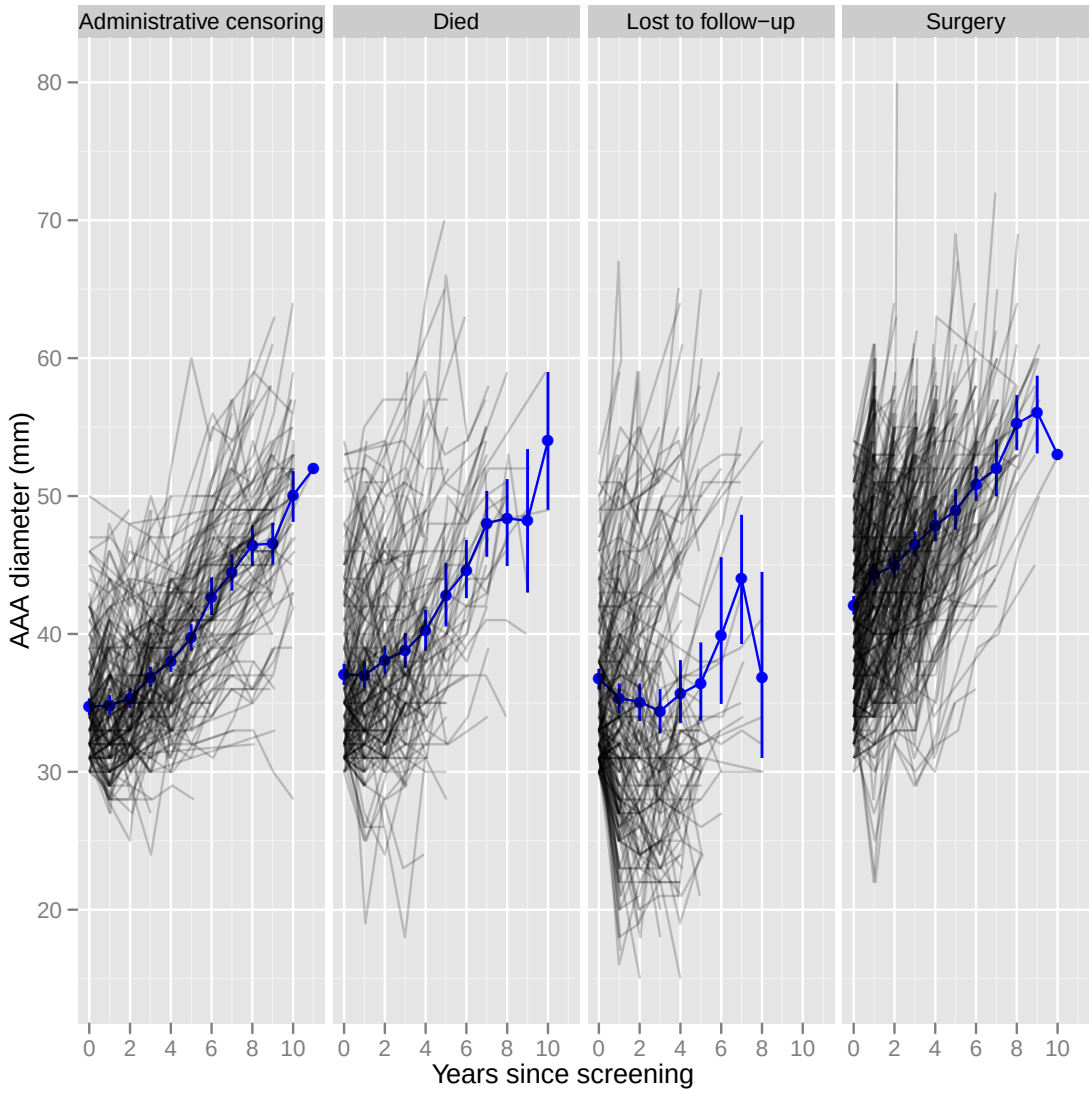

Figure 2: Trajectories of AAA growth given the type of censoring for all yearly AAA observations. The yearly mean AAA diameter with 95% nonparametric bootstrap confidence intervals are superimposed on the plots.

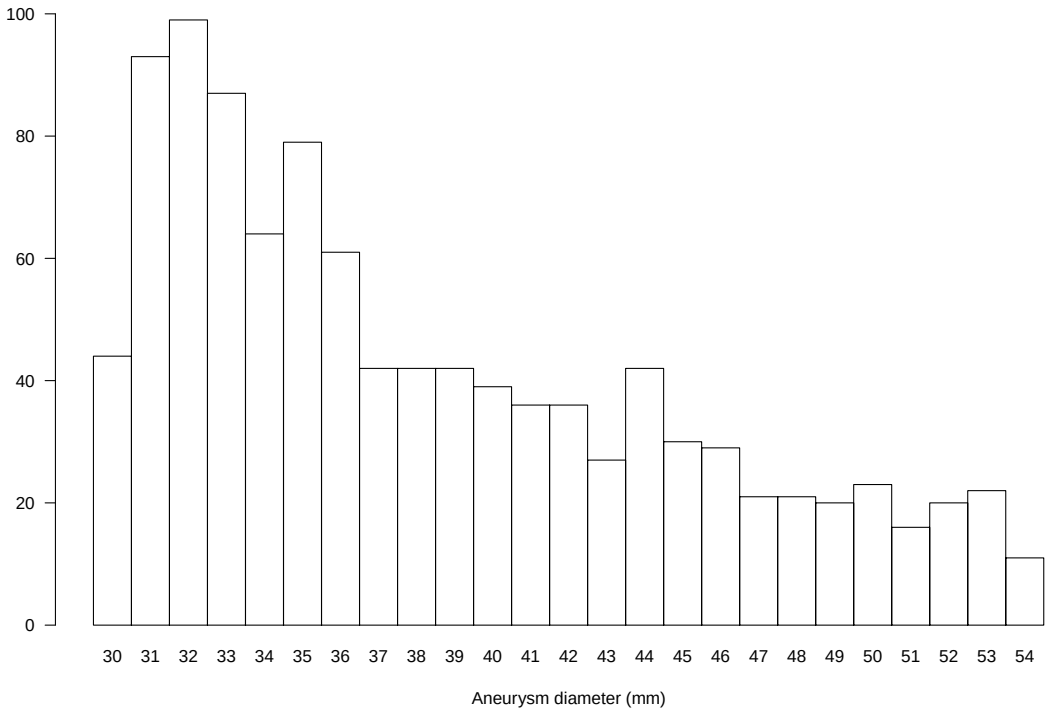

Figure 3: Histogram of aneurysm diameters at first screen.

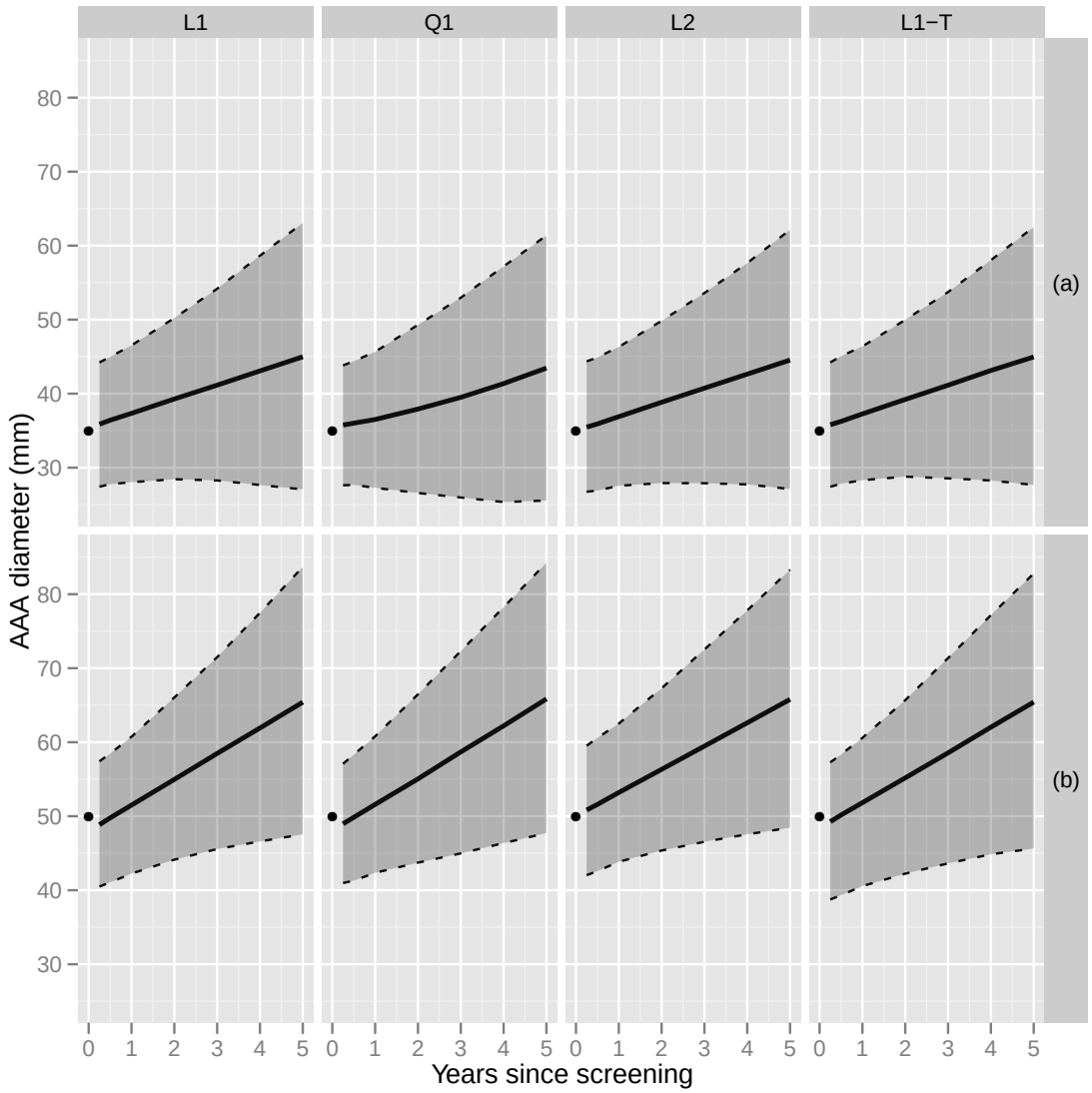

Figure 4: Predicted AAA diameter given a current diameter of either (a) 35mm or (b) 50mm taken at time of screening. Posterior medians and 95% credible intervals are presented.

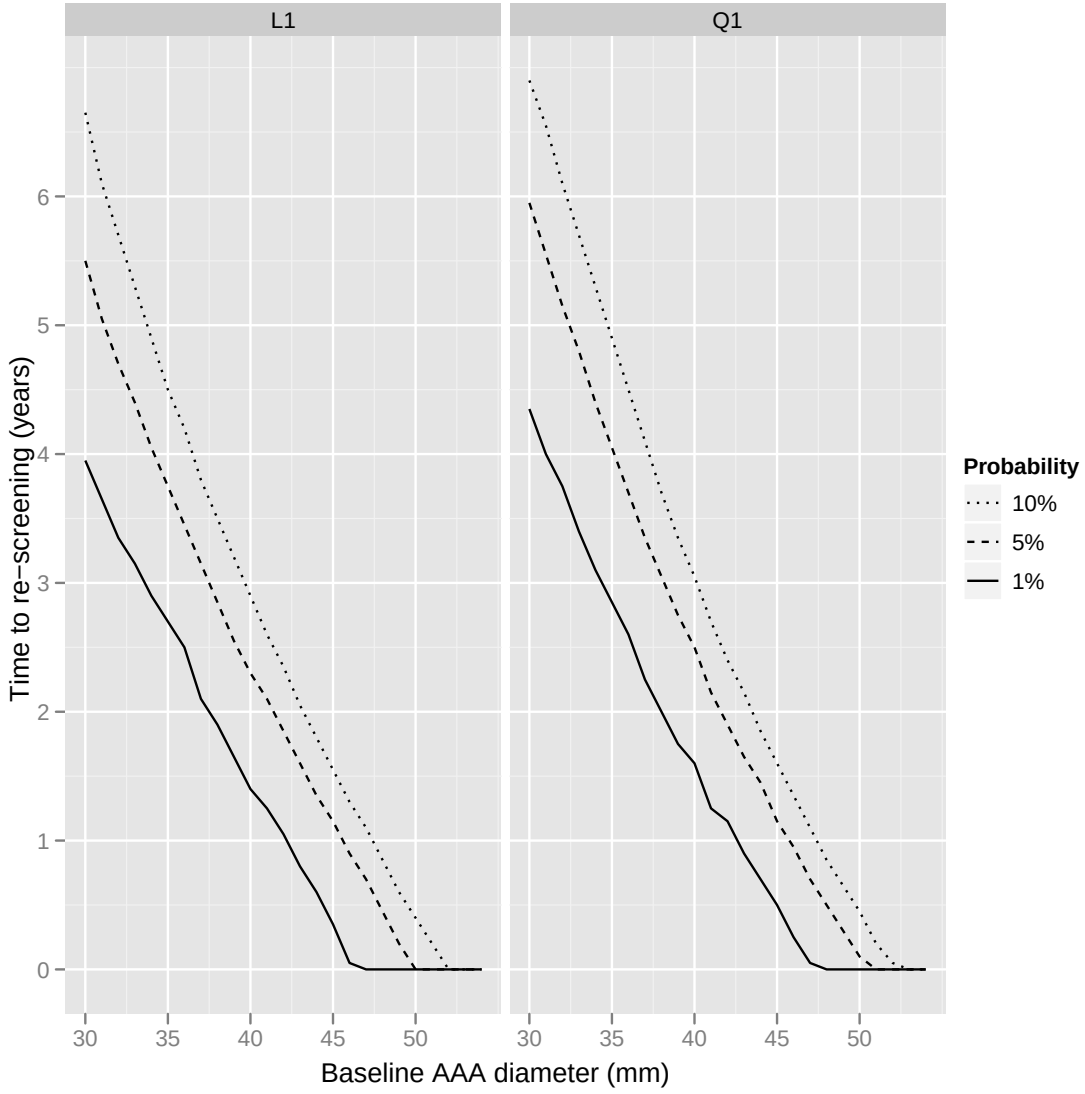

Figure 5: Probability of an observed AAA diameter exceeding 55mm at re-screening given baseline AAA diameter.

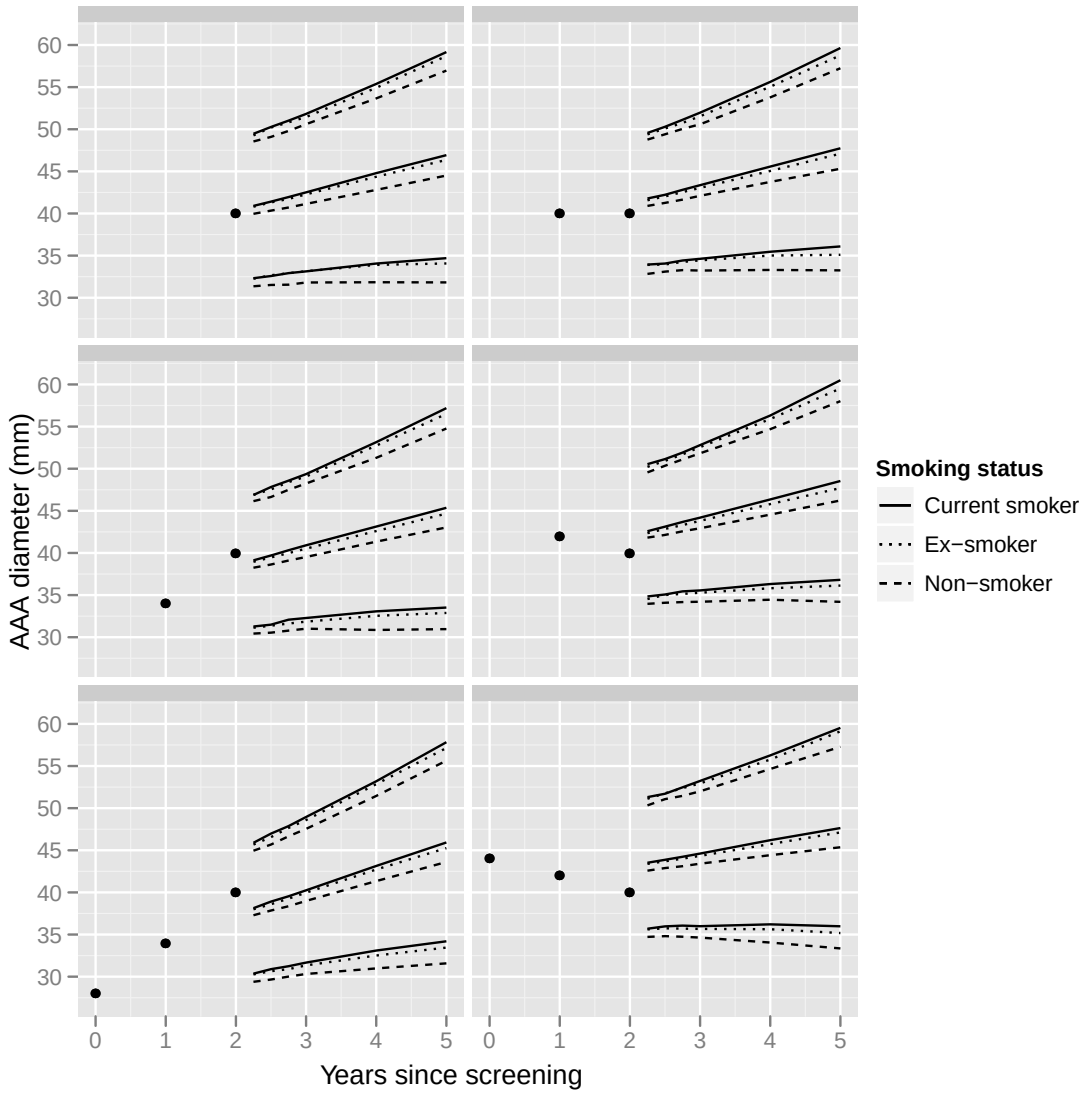

Figure 6: Predicted AAA diameter for an individual aged 65 at screening, by smoking status and observed growth. Posterior medians and 95% credible intervals are presented for each prediction.