Research Article

Statistics in Medicine

Received 21 January 2009, Accepted 6 May 2010 Published online 18 June 2010 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.3992

Analysis of multi-arm tumor growth trials in xenograft animals using phase change adaptive piecewise quadratic models

Kingshuk Roy Choudhury[a*†], Ian Kasman[b] and Greg D. Plowman[b]

Xenograft trials allow tumor growth in human cell lines to be monitored over time in a mouse model. We consider the problem of inferring the effect of treatment combinations on tumor growth. A piecewise quadratic model with flexible phase change locations is proposed to model the effect of change in therapy over time. Each piece represents a growth phase, with phase changes in response to change in treatment. Piecewise slopes represent phase-specific (log) linear growth rates and curvature parameters represent departure from linear growth. Trial data are analyzed in two stages: (i) subject-specific curve fitting (ii) analysis of slope and curvature estimates across subjects. A least-squares approach with penalty for phase change point location is proposed for curve fitting. In simulation studies, the method is shown to give consistent estimates of slope and curvature parameters under independent and AR (1) measurement error. The piecewise quadratic model is shown to give excellent fit (median $R^2 = 0.98$) to growth data from a six armed xenograft trial on a lung carcinoma cell line. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: tumor growth modeling; piecewise quadratic model; penalized least squares

1. Introduction

Developments in basic biology and biotechnology have led to a large number of new potential therapies for treatment of cancer [1]. Tumor growth experiments are a convenient way to test the efficacy of potential new therapies for various forms of human cancer in mice. In a tumor growth experiment, *in vitro* cultured human tumor tissue (xenograft) is implanted subcutaneously into immunosuppressed mice. Experimenters monitor the growth of these tissue fragments into tumors of measurable size, which happens approximately 15 days after implantation. Measurements can be made once they are of palpable size (approximately $100\,mm^3$), i.e. they can be measured externally by pinching the skin over the tumor mass and recording the tumor length and width using calipers. Once measurability is established, mice are randomized to various treatment groups and treatment is started at a common date. Tumor size (usually quantified as tumor volume) for these mice is recorded regularly at evenly spaced time points till the end of the experiment (usually a few weeks). Each mouse is euthanized at the end of the study, or at the first time point after its tumor size exceeds a preset threshold V_{max} (this last observation is recorded). Treatments consist of placebos, single drugs, drug combinations [2] or other therapy, such as surgery or radiation [3]. Typically, it is of interest to model tumor growth during treatment, i.e. the period when drug is being delivered and identify the most effective treatments, as well as any 'synergy' when treatments are offered in combination. It is also of interest to study whether the rate of tumor growth after drug(s) has been turned off, known as the 'regrowth' or 'rebound' period.

[a]*Statistics Department, University College Cork, Cork, Ireland*
[b]*Department of Tumor Biology and Angiogenesis, Genentech, Inc, South San Francisco, CA, U.S.A.*
[*]*Correspondence to: Kingshuk Roy Choudhury, Statistics Department, University College Cork, Cork, Ireland.*
[†]*E-mail: kingshuk@ucc.ie*

Contract/grant sponsor: Science Foundation Ireland; contract/grant number: MATF543/2007

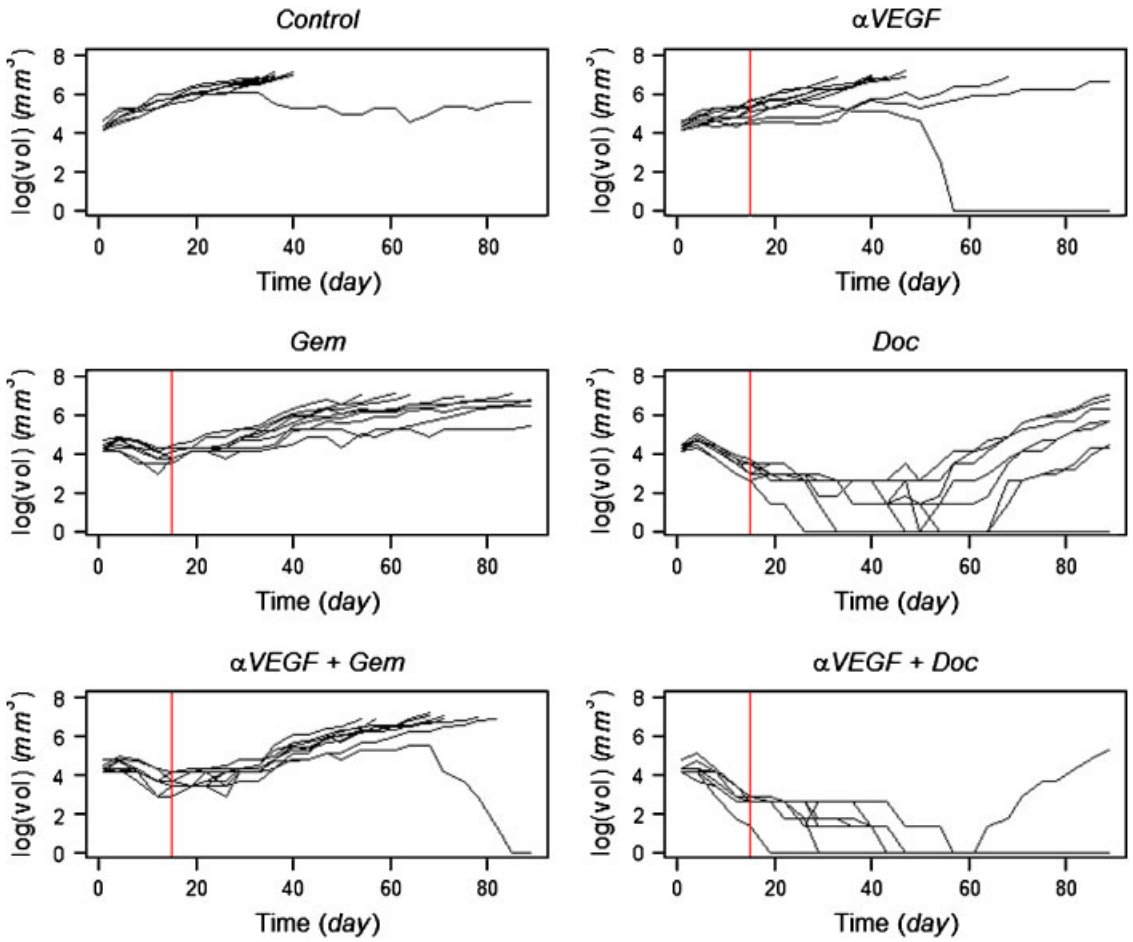

Figure 1. Plots of tumor volume growth curves for a six armed study involving MV-522 (lung carcinoma) tumor xenografts against time (Section 4.1). The study lasted 89 days. Animals whose tumor volumes exceeded 1000 mm^3 were euthanized. Details of treatments and dosing regimens are given in Table I. Each line corresponds to a single animal. Zero volumes are imputed as 0 on log scale. Day 1 represents the first day of treatment, approximately 2 weeks after implantation of the tumors. Vertical lines indicate the last treatment dose.

1.1. Statistical modeling of tumor growth experiments

Xenograft studies produce longitudinal data consisting of a set of tumor size measurements $V_{gj}(t)$, for animals $j=1,\ldots,n_g$ in treatment groups $g=1,\ldots,G$ and time points $t=t_0, t_1,\ldots,t_m$. Here t_0 is the first date of therapy administration and t_m is the study end. For animal j, data are only recorded till time $e_{gj}=\min\{l: V_{gj}(t_l)>V_{\max}\}$, where $V_{\max}$ is a preset size threshold. With longitudinal data, we need to model two types of variability: (i) variation in time (t) (ii) variation across subjects (j) and treatment arms (g). A basic model for tumor growth over time, based on repeated cell division, is the exponential model:

$$V(t)=V(0)\exp(\lambda t). \tag{1}$$

Here $V(t)$ denotes tumor volume at time t and λ denotes (constant) growth rate. This model has been shown to be effective when modeling data for a relatively short period of time during the fast growth period of a tumor, as demonstrated by Skipper and Schabel [4]. During this phase of growth, the exponential model of growth is motivated by binary cell division with constant doubling time [5]. In Figure 1, we see that tumors in the control and αVEGF arms appear to exhibit roughly log-linear growth (except possibly for a couple of animals), which is consistent with the Skipper model (1). However, the exponential model fails to adequately describe the more complex tumor growth patterns observed in some of the other arms of the experimental data shown in Figure 1. More sophisticated models have been proposed to model the effect of treatment, e.g. a mixture double exponential model with rates for cell growth, cell death and surviving cell fraction [6]. A four-parameter model of tumor growth based on compartmental modeling was proposed in [7]. This model also accommodates time-dependent dosing.

The methods described above have all been shown to be successful for modeling data from tumor growth experiments. In addition, they have scientific interpretability in terms of representing specific physiological processes such as cell division, cell death, clearance of dead cells, etc. [3, 6]. However, specific models of growth can be a limitation when making comparisons across a variety of different tumor types and treatment combinations, whose mechanism of action may be different or indeed, unknown [3]. Additionally, different models will not necessarily be nested, making parameter-based comparisons difficult. It is therefore desirable that these models: (i) have common functional form across treatments and tumor types, enabling parameter comparisons across these factors, (ii) fit the data well, (iii) have parsimonious parameterization, (iv) incorporate dosing information and (v) can be fit to data using automatic methods. In Section 2, we present a piecewise quadratic approach to modeling tumor growth data. Properties of resulting estimators are examined using a set of simulation studies in Section 3. The methodology is illustrated by application to a multi-armed experiment in Section 4.

2. Tumor growth curve modeling

2.1. Piecewise quadratic model

Our modeling paradigm is based on the empirical observation, across a number of studies that tumor growth appears to occur in distinct phases: e.g. during treatment and after treatment. This type of behavior is illustrated through a six armed study (with one control and five treatment groups) described in Section 4 and visualized through longitudinal profiles in Figure 1. In Figure 1, all treatments groups appear to exhibit a two-phase growth regime, except for the control and αVEGF groups, which appear to exhibit a one-phase growth regime. To capture tumor growth in different phases, we propose a subject-specific piecewise model:

$$Y_{gj}(t) = \mathrm{E}_{\theta_{gj}}[Y_{gj}(t)] + \varepsilon_{gj}(t)$$
$$\text{where } E_{\theta_{gj}}[Y_{gj}(t)] = \sum_{k=1}^{K_g} P(t, \theta_{gjk}) I[\tau_{gj(k-1)} \leqslant t < \tau_{gjk}]. \tag{2}$$

Here, $Y_{gj}(t) = \log V_{gj}(t)$ and $P(t, \theta_{gjk})$, $k = 1, \ldots, K_g$ are pieces modeling the behavior of Y_{gj} on the interval $\tau_{gj(k-1)}$ to τ_{gjk}, with $\tau_{gj0} = t_0$, $\tau_{gjk} = t_m$, $\tau_{gj(k-1)} < \tau_{gjk}$ and $\theta_{gj} = \{\theta_{gjk}, k = 1, \ldots, K_g\}$. The expectation in (2) is taken with respect to the distribution of measurement error, $\varepsilon_{gj}(t)$, i.e. within subject variation. Each piece represents one phase of the treatment regime. The number of phases depends on the sequence of treatments applied to the subject. In Figure 1, $K_g = 1$ for the control group (no treatment). For all other treatment arms, $K_g = 2$, with one piece representing the treatment period and the second piece representing the regrowth/recovery period. Phase change points τ_{gj1} do not necessarily coincide with treatment change dates. Drug combinations administered simultaneously count as a single treatment.

The simplest parametric model for growth is piecewise linear: $P(t, \theta_{gjk}) = a_{gjk} + b_{gjk} t$, where $\theta_{gjk} = (a_{gjk}, b_{gjk})$. Here a_{jk} represents $\log(V_{gj}(\tau_{gj(k-1)}))$, the initial tumor volume at commencement of the kth phase and b_{gjk} represents the (linear) growth rate for the kth phase. In Figure 1, a linear model appears to be adequate for the control group. However, for the Doc treatment group, many growth profiles exhibit a trough shape, which is difficult to approximate well by two linear pieces. To better accommodate this, we propose a piecewise quadratic model: $P(t, \theta_{gjk}) = a_{gjk} + b_{gjk} t + c_{gjk} t^2$, where $\theta_{gjk} = (a_{gjk}, b_{gjk}, c_{gjk})$. Here c_{gjk} determines the curvature or non-linearity of growth: $c_{gjk} = 0$ represents linear growth, $c_{gjk} > 0$ denotes increasing and $c_{gjk} < 0$ denotes decreasing growth rate. Since the question of non-linearity is of interest in any phase of growth, we will apply the piecewise quadratic model to all subjects across all groups. Thus in subsequent analysis, the parameter set θ_{gjk} is always three dimensional. With this formulation, it is possible to compare parameter sets across groups and phases, e.g. b_{1j1}, the growth rate for the control group, which has only one phase, can be compared with b_{2j2} the growth rate for the recovery (second) phase of treatment group 2. Comparison of treatment groups is accomplished by a second-stage analysis of the subject-specific parameter sets θ_{gjk}, also known as 'derived variables', across subjects and treatments: details are given in Section 2.5.

2.2. Measurement error model

Assuming the tumor to be a prolate spheroid, i.e. cigar shaped, the volume of a tumor is $V = \pi w^2 l/6$, where l and w are the largest and smallest diameters of the tumor, respectively, measured using calipers [8]. Assuming additive Gaussian errors, we can model the measurements as:

$$w = w_0 + \varepsilon_w = w_0(1 + w_0^{-1}\varepsilon_w); \quad l = l_0 + \varepsilon_l = l_0(1 + l_0^{-1}\varepsilon_l). \tag{3}$$

Here, l_0 and w_0 are the actual diameters and ε_w and ε_l are $N(0, \sigma^2)$ errors in measurement. We can therefore write $\log V$ as:

$$\log V = \log \varsigma + 2\log(1 + w_0^{-1}\varepsilon_w) + \log(1 + l_0^{-1}\varepsilon_l) \approx \log \varsigma + \varepsilon_v. \tag{4}$$

Here, $\varsigma=\pi w_0^2 l_0/6$ is the true volume and $\varepsilon_v=2w_0^{-1}\varepsilon_w+l_0^{-1}\varepsilon_l$. Equation (4) suggests that $Y=\log V$ has an approximately additive error structure, because if ε_w and ε_l have zero mean, then so has ε_v. Thus, log is a natural scale to model tumor volume data. When tumor sizes are small, the log transformation can, however, lead to erroneous results [9].

2.3. Fitting piecewise models by penalized least squares

We propose a subject-specific weighted least-squares objective function to fit the piecewise tumor growth model as follows:

$$L_{gj}=\sum_{i=1}^{e_{jg}}\left(Y_{gj}(t_i)-\sum_{k=1}^{K_g}P(t,\theta_{gjk})I[\tau_{gj(k-1)}\leqslant t_i<\tau_{gjk}]\right)^2. \tag{5}$$

The piecewise nature of the model (2) allows the least-squares criterion to be broken into K_g sub-criteria: $L_{gj}=\sum_{k=1}^{K_g}L_{gjk}$, where $L_{gjk}=\sum_{t\in S_{gjk}}(Y_{gj}(t)-P(t,\theta_{gjk}))^2$ and $S_{gjk}=\{t_l:\tau_{gj(k-1)}\leqslant t_l<\tau_{gjk}\}$. Given the sets S_{gjk}, we can minimize the criterion L_{gj} by minimizing each L_{gjk} separately. For two-phase models, we include the first data point of the second phase in the criterion for the first phase, L_{gj1}. We have empirically observed that this yields approximate continuity of the fitted growth curve. Minimization of each sub criterion can be achieved through ordinary least-squares (OLS) regression. Additionally, we center the time variable using the transformation $s=t-\bar{t}_{gjk}$, where $\bar{t}_{gjk}$ is the mean of t in S_{gjk}. Thus, we rewrite $P(t,\theta_{gjk})=a_{gjk}+b_{gjk}s+c_{gjk}s^2$ and the observation as $Y_{gj}(s)$. Although identical fits will be obtained with or without this transformation, the variables s and s^2 as well as 1 and s are approximately orthogonal. In this case, estimates of the intercept a_{gjk} slope b_{gjk} and curvature parameter c_{gjk} become (approximately) mutually uncorrelated. Hence separate inference can be performed on these parameters, see e.g. example 15.1 in [10]. Furthermore, this transformation ensures that the estimator of slope b_{gjk} will remain approximately the same if we instead used just a piecewise linear approximating function, $P(t,\theta_{gjk})=a_{gjk}+b_{gjk}t$, with or without centering.

In two-phase models, where there is a clear change in growth rate, the phase change point can be identified by minimizing the criterion, L, over all admissible phase change points τ_{gj1}. However, in the special case when treatment change has no effect, i.e. $b_{gj1}=b_{gj2}$, $c_{gj1}=c_{gj2}$, it is not possible to identify a unique phase change point using only the least-squares criterion L_{gj}, since any choice of point between τ_{gj0} and τ_{gj2} will yield the same value of L_{gj}. This scenario can occur for instance when placebo is used. In practice, noise in the data may lead to a unique minimum, but the resulting phase change point may be at an extremal location, leading to potentially highly variable parameter estimates. To ensure the choice of more meaningful phase change points, we propose to penalize phase change location by

$$\begin{aligned}D_{gj}(\tau_{gj1})&=\infty \quad \text{if} \quad \tau_{gj1}<T_{g1}\\ &=v_{g1}^{-1}(\tau_{gj1}-T_{g1})^2 \quad \text{if} \quad \tau_{gj1}\geqslant T_{g1},\end{aligned} \tag{6}$$

where T_{g1} is the end time of the first treatment period, v_{g1} is the standard deviation of phase change points for this treatment. The asymmetric nature of D_{gj} reflects our focus on measuring the effects of treatment change: change in growth pattern before change in treatments would not reflect a causal effect, hence it is excluded. We combine (5) and (6) to obtain a penalized least-squares criterion: $L_{gjP}=L_{gj1}+L_{gj2}+\lambda D_{gj}$, where λ is a parameter that controls the influence of the penalty term. As $\lambda\to\infty$, the criterion L_{gjp} is dominated by the penalty term, which has a unique minimum at T_{g1}. On the other hand, small values of λ will lead to a solution that is close to least squares. We recommend choosing the smallest λ which ensures a unique minimum for L_{gjP} for all subjects. Prior information about the standard deviation v_{g1} is typically unavailable for experimental drugs and drug combinations. Instead, we have used an empirical estimate $\hat{v}_{g1}^2=(n_g-1)^{-1}\sum(\hat{\tau}_{gj1}-T_{g1})^2$, the sample variance of phase change points $\hat{\tau}_{gj1}$, each estimated by visual inspection, across subjects within group g. A scaled version of the penalized criterion L_{gjP}, namely $G_{gjP}=1-L_{gjP}/\text{Var}(Y_{gj})$, where Var stands for sample variance, is used to choose the optimal phase change location (Figure 3(b)). Note that G_{gjP} may be negative.

2.4. Complete fitting algorithm

The main steps in fitting the piecewise quadratic model can be summarized as follows:

1. Determine range of admissible phase change locations from dosing regimen. Fix $\lambda=0.0001$.
2. Choose a candidate phase change point p from feasible set.

(a) Fit piecewise quadratic model:

(i) Compute transformed observation times for each piece
(ii) Fit piecewise parametric model by least squares

(b) Compute G_{gjP} value for fitted piecewise quadratic model.

3. Maximize G_{gjP} over all admissible change points p. If maximum is not unique, choose a bigger λ (e.g. twice old λ) and repeat from step 2.

The above algorithm can be easily implemented on any statistical software package that will fit linear regression and also allows looping, e.g. R, SAS, Stata, etc.

2.5. *Analysis across subjects*

The above fitting procedure generates subject-specific parameter sets $\theta_{gjk}=(a_{gjk}, b_{gjk}, c_{gjk})$ for animals $j=1,\ldots,n_g$ in treatment groups $g=1,\ldots,G$ and treatment phases $k=1,\ldots,K_g$, where $K_g=1$ for the control group, while $K_g=2$ for all other treatment arms. Under the derived variables approach, these estimated parameter sets are used as data points for the second stage of analysis, across subjects and treatment groups [11]. For analysis of growth, the specific parameters of interest are slopes, b_{gjk} and curvatures, c_{gjk}. We have argued in Section 2.3 that estimates of these parameters are approximately mutually uncorrelated. Furthermore, since θ_{gj1} and θ_{gj2} are calculated using data from mutually separate time points (except one common endpoint), under the assumption of independent measurement error, estimates of these parameter sets will also be approximately mutually uncorrelated. This mutual uncorrelatedness allows separate analysis of slopes and curvatures within each phase of treatment, considerably simplifying the second-stage analysis across subjects, groups and phases. For instance, the analysis of first (treatment) phase slopes can be done in the framework of a one-way ANOVA model, across groups $g=1,\ldots,G$, using the estimated b_{gj1} as data points. Thus, hypotheses about differences in mean growth rates between groups can be addressed using a standard linear model framework. For the analysis of second- (recovery) phase slopes, we add estimates for the first-phase slopes of the control group, b_{gj1} to the rest of the second-phase slopes b_{gj2}, $g=2,\ldots,G$. Again, analysis of this combined set of slopes under the linear model framework enables us to test hypotheses of differences between recovery (second) phase and control growth rates. Finally, the analysis of curvatures c_{gjk} is performed within each group g and phase k, with the null hypothesis that the phase-specific group mean curvature is zero, i.e. growth is linear for that particular phase of treatment.

3. Consistency of parameter estimates

In this section, we study properties of the slope and curvature parameter estimates obtained by fitting the piecewise quadratic model. Standard results for parameter estimates in linear models do not necessarily apply, because (a) of volume endpoint censoring, which means that the number of observations per subject is data dependent, (b) adaptive phase change point selection in the piecewise model (2), (c) randomness of the subject-specific initial tumor volume and (d) potential dependence in the longitudinal tumor measurements [11]. In addition, the piecewise model is not differentiable, making analysis of the estimator difficult (see Discussion). Using simulation studies, we examine whether estimates of the growth rate and curvature parameter are consistent for true parameter values and remain robust under settings that replicate these four sources of variation, as well as changes in the growth rate and level of measurement error.

In the simulation study, we consider a two-phase tumor growth model. Tumor volumes are recorded every 3 days for a period of 90 days. In keeping with the sampling scheme for xenograft experiments, tumor volumes are measured until the first time the tumor volume exceeds $V_{\max}=1000\,\text{mm}^3$. Data are generated as:

$$y_i = a + b_1 t_i + (b_2 - b_1)(t_i - 23)_+ + \sigma\varepsilon_i. \tag{7}$$

Here y_i, $i=0,\ldots,29$, is the log tumor volume recorded at time t_i and $\sigma\varepsilon_i$ is the measurement error. A (potential) change in the growth rate occurs on day 23, the direction of change (increase or decrease) depending on the relative values of b_1 and b_2. The change follows end of treatment on day 18 ($i=6$). For this model, data were generated by choosing the intercept (initial log tumor volume), with $\exp(a)$ drawn as a random sample from a $N(50\,\text{mm}^3, 10\,\text{mm}^3)$ distribution. A range of 5 equispaced values between 0.05 and 0.15 were used for the initial growth rate b_1 and a range of 7 equispaced values between 0.05 and 0.2 were used for second-phase growth rate b_2. Data were generated for a range of 100 values of standard deviation σ between 0.1 and 0.5, equally spaced on a logarithmic scale. The parameters b_1, b_2 and σ are dimensionless, because $y_i - a = \log(V(t)/V(0))$ is dimensionless. For each realization, the squared error of estimation was computed as: $\text{SqE}(\hat{b}_k, \sigma)=(\hat{b}_k - b_k)^2$, $\text{SqE}(\hat{q}_k, \sigma)=(\hat{q}_k - q_k)^2$, $k = 1, 2$, where $\hat{b}_k$ and $\hat{q}_k$ are estimated slope and

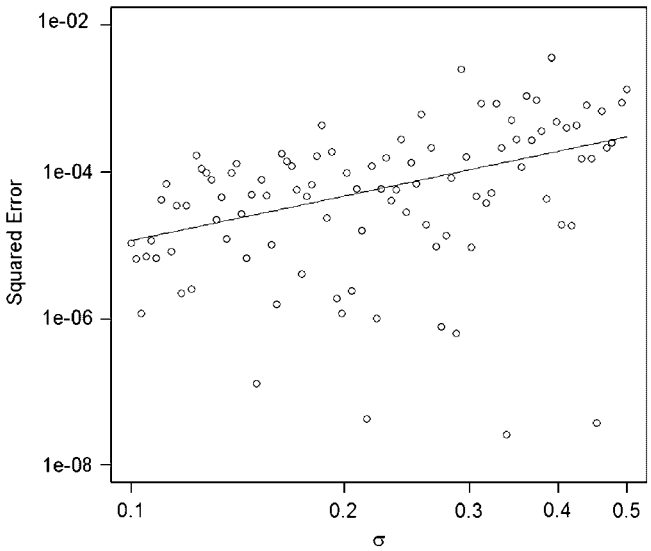

Figure 2. Results of one run of the simulation experiment. Squared estimation error for the slope of the first piece of a two piece linear growth model given in (7), with $b_1 = 0.083$, $b_2 = 0.20$, for a range of 100 different noise levels σ equispaced between 0.1 to 0.5, plotted on a log–log scale. The solid line shows the fitted mean-squared error as a log-linear function of σ. Fitting was done using the robust (M) regression.

curvature parameters obtained by fitting a piecewise quadratic model of the form (2), with $P(t, \theta_k) = a_k + b_k t + q_k t^2$, $k = 1, 2$, using the procedure described in Section 2.3. For each data set, the change point τ_1 was estimated over a range of time points starting from day 18 (end of treatment) to the last observed timepoint. Only 1 replication per parameter set (b_1, b_2 and σ) was carried out for a total of: $5 \times 7 \times 100 = 3500$ parameter sets. Results are now presented for two different specifications of measurement error.

IID case: Here, we assume $\varepsilon_i \sim$ i.i.d. $N(0, x)$. In Figure 2, we see that the mean-squared error for estimation of $b_1$increases linearly (on a log–log scale) with σ for one set of b_1, b_2 values. The rate of decrease/convergence is estimated by fitting a line of the form: $\log \text{SqE}(\hat{b}_1, \sigma) = m_b + r \log \sigma$ to the squared error data using robust (M) regression [12], due to the presence of outliers at the lower end of the plot. Similar lines were fit for all 35 b_1, b_2 sets in the simulation. The average slope obtained across these parameter sets was $\bar{r} = 1.94$ with a standard deviation of 0.37. This indicates that $\hat{b}_1$ converges to the true parameter value b_1 at the usual parametric rate [13]. Corresponding average convergence rates and standard deviations (in parentheses) obtained for the estimates, $\hat{b}_2, \hat{q}_1$, and $\hat{q}_2$ were 2.50 (0.56), 1.97 (0.50), 3.04 (0.81), indicating that these estimators too converged at the desired parametric rate.

Correlated errors: Here, we assume that the errors are realizations of an AR(1) process (see Section 3.2). Specifically, $\varepsilon_i = \varphi \varepsilon_{i-1} + \eta_i$, where $\eta_i \sim$ i.i.d. $N(0, x)$, with $\phi = 0.2$. Other simulation settings are similar to above. Note that estimates were computed using the OLS criterion (5), i.e. no allowance was made for correlation between observations while fitting the models. As previously, we computed the average rate of convergence across 35 b_1, b_2 parameter sets in the simulation. The average rates of convergence and standard deviations (in parentheses) obtained for the estimates, $\hat{b}_1, \hat{b}_2, \hat{q}_1$ and $\hat{q}_2$ were 2.05 (0.39), 2.15 (0.67), 1.89 (0.46), 2.35(0.85) indicating that the OLS estimators converged to the true parameter values at the desired parametric rate. Furthermore, we also computed parameter estimates using generalized least squares (GLS) for each simulation realization, where the weighting matrix was the inverse of the variance covariance matrix of the errors, computed using the true value of $\phi = 0.2$. Although this is not a practical option (correlation will have to be estimated in practice, see Discussion), with correlated errors, we would expect this estimator to be optimal. We compared the OLS and GLS-squared errors using a paired t-test across noise levels σ. For all 35 b_1, b_2 parameter sets, the resulting p-values were not significant, i.e. above 0.05.

4. Analysis of multi-armed tumor growth experiment

4.1. Experimental protocol

Xenografts were initiated from MV-522 human non-small cell lung carcinoma tumors maintained by serial transplantation in female nude (*nu/nu*, Harlan) mice. Each test mouse received a subcutaneous MV-522 tumor fragment ($1\,\text{mm}^3$) implanted in the right flank. Nude mice are athymic due to a genetic mutation. Because they lack a thymus gland, they

Table I. Protocol design for the six armed MV-522 xenograft study.

Protocol design							
Group	n	Drug 1	Dose (mg/kg)	Duration	Drug 2	Dose (mg/kg)	Duration
1	9	Control					
2	9	αVEGF	10	1/week × 3			
3	9	Control			Gem	160	1/4 days × 3
4	9	Control			Doc	30	1/week × 3
5	9	αVEGF	10	1/week × 3	Gem	160	1/4 days × 3
6	9	αVEGF	10	1/week × 3	Doc	30	1/week × 3

The two factors are drug 1 (the presence or absence of anti-angiogenic factor αVEGF) and drug 2 (no drug, gemcitabine or docetaxcel). Dosing regimen describes how often and how long each drug is administered, e.g. 1/week × 3 means: 1 dose per week for 3 weeks. n stands for number of mice per treatment arm.

are deficient in T cells and so do not present an immune rejection response against the injected tumor cells, which allows a high proportion of tumor fragments to grow into fully fledged tumors [2]. The growth of tumors was monitored as the average size approached 80–120 mm^3. Fifteen days later, designated as Day 1 of the study, the animals were sorted into six size matched groups (each with $n=9$) with individual tumor volumes between 63 and 12 mm^3 and group mean volumes of 75 mm^3. All tumors were measured twice weekly (on the same day) using calipers (Figure 1). Animals were euthanized when their tumor size exceeded 1000 mm^3 or at the end of the study (Day 89), whichever came first. An Institutional Animal Care and Use Committee approved all animal protocols.

4.1.1. Treatments and study design. A two factor experiment was performed. The design is given in Table I. One factor is the presence or absence of chemotherapy (gemcitabine (Gem) or docetaxel (Doc)). Gemcitabine (Gemzar®, Eli Lilly and Company) and docetaxel (Taxotere®, Aventis Pharmaceuticals) were administered and evaluated at the maximum tolerated dose (MTD), defined as the dose level immediately above which excessive toxicity (i.e. more than one death) occurred. The other factor is the presence or absence of anti-vascular endothelial growth factor (αVEGF), which is a murine IgG2a monoclonal antibody (B20-4.1) from Genentech [14] that recognizes all isoforms of VEGF-A with comparable affinity for both the mouse and human ligands. A murine IgG2a Mab specific to ragweed was used as a control. Previous pharmacokinetic studies support their use at 10 mg/kg once a week. When administered on the same day, chemotherapy drugs were administered within 30 min after antibodydosing.

4.2. Analysis using piecewise quadratic fits

In Figure 1, tumors in the control and αVEGF arms appear to exhibit roughly log-linear growth. For all other arms, we see that the growth curves appear to be qualitatively affected by the treatment. Two-piece quadratic models ($K_g=2$) were fit to all the curves in the MV-522 study using the penalized least-squares algorithm described in Section 2.3 (with $\lambda=0.001$), except curves in the control arm, where a one-piece model was used ($K_g=1$). The values of the v_{g1} parameter in the penalty term were estimated to be 4 in the αVEGF, Gem and (Gem+αVEGF) treatment arms and 10 in the Doc and (Doc+αVEGF) treatment arms.

An example of this fitting criterion is shown in Figure 3 for a subject in the Doc treatment arm from Figure 1. In this case $K_g=2$ and $T_{g1}=15$. A value of $\lambda=0.001$ was chosen by grid search. The algorithm searched over a range of phase change point candidates τ_{gj1}, from 15 to 60. The median G_{gjP} value across treatment arms was 0.98 with a mean absolute deviation of 0.01, indicating very good fit in general, although there were one or two exceptions. Residuals in the control arm were found to have sample autocorrelations of $\mathrm{Corr}(\eta_t,\eta_{t-1})=0.23$, $\mathrm{Corr}(\eta_t,\eta_{t-2})=0.1$ and $\mathrm{Corr}(\eta_t,\eta_{t-3})=0.01$. The decay in autocorrelations suggests an AR(1) model with $\phi=0.2$ for intrasubject correlation [11]. The quantile–quantile analysis of residuals showed them to be approximately Gaussian. Autocorrelation analysis of residuals in other arms was not pursued because of potential anomalies due to incorrect positioning of phase change points.

Estimated phase change locations for the Gem and (Gem+αVEGF) treatment arm appear to be broadly similar across subjects and are generally close to the end of treatment (Figure 4(a)). By contrast, phase change locations for the Doc and (Doc+αVEGF) treatment arm have relatively higher variability, with many locations indicating a long delay for regrowth after treatment. The phase change locations for the αVEGF treatment arm are somewhat artificial, induced by the penalty term in the fitting criterion. This is confirmed by the fact that there are no significant differences in the slope (p-value=0.34) and curvature parameter (p-value=0.15) estimates for the two pieces fitted to this treatment arm under paired t-tests.

The distribution of slope parameter estimates exhibits substantial separation for certain treatment arms relative to control (Figure 4 (b)). The slopes in the control arm mostly lie between 0.05 and 0.1. Treatment phase growth rates (b_1)

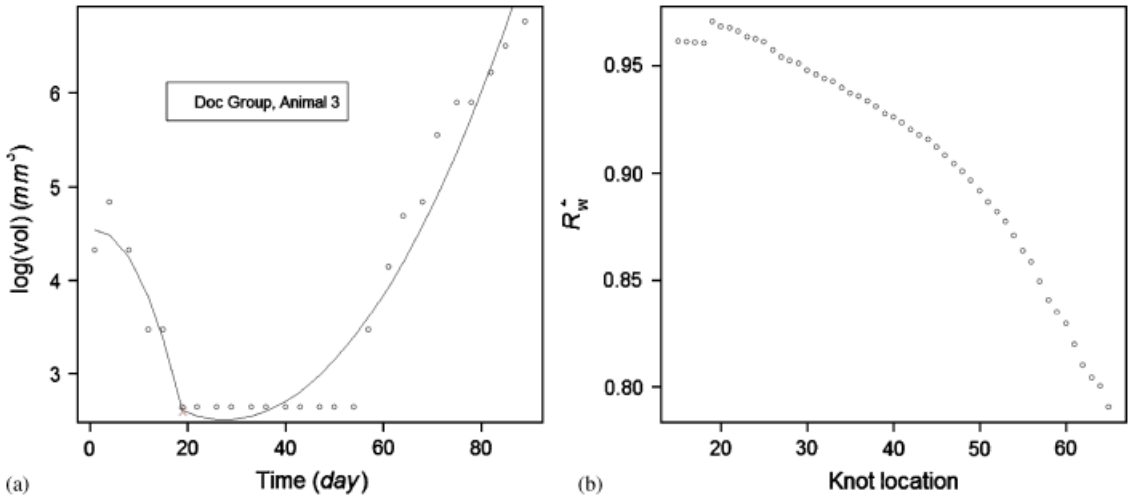

Figure 3. Example of piecewise quadratic fit to a growth curve from the docetaxel arm: (a) tumor volume data for one animal and corresponding fit. Estimated phase change location is at 19 days; (b) plot of G_{gjP} statistic used to choose phase change location. The penalty function L_p was computed with standard deviation of change points $(v_{g1})=10$ days and $\lambda=0.001$.

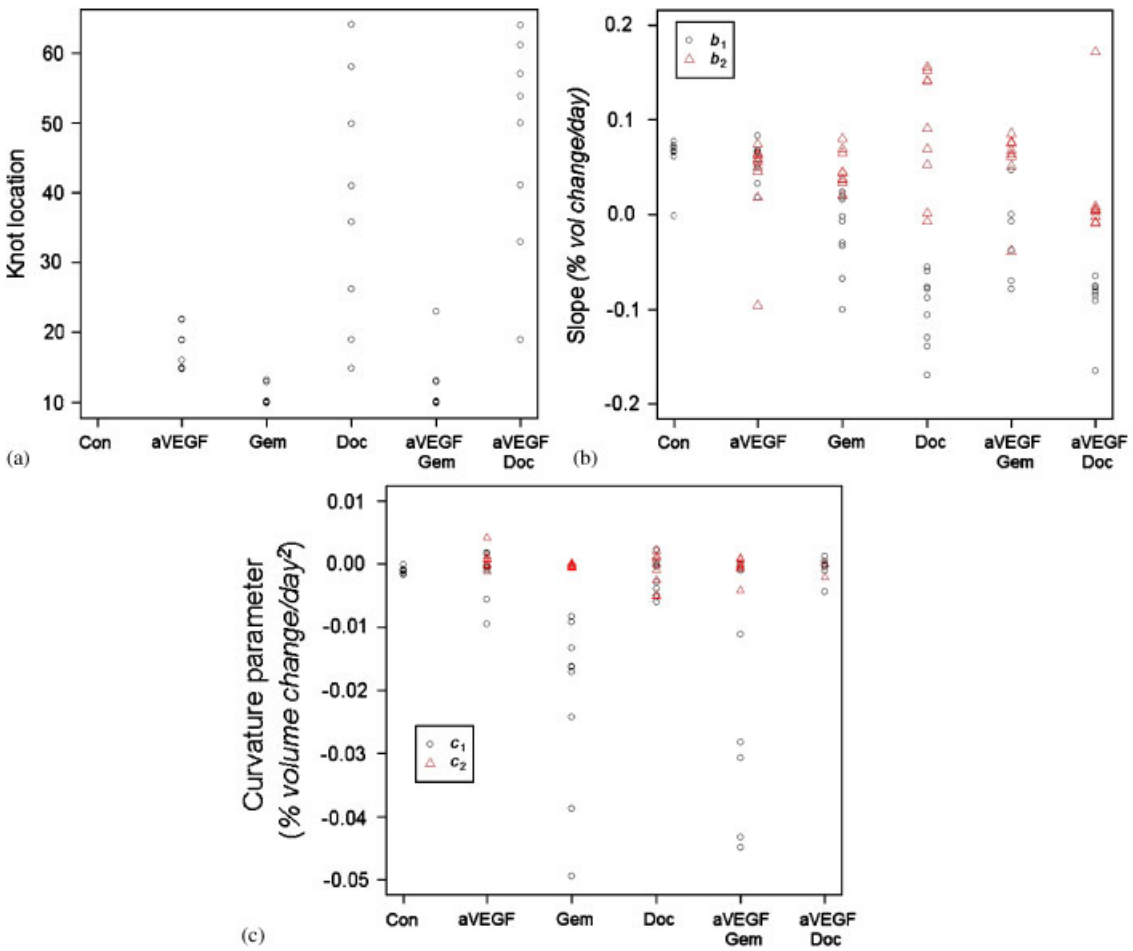

Figure 4. Between subjects analysis for the MV-522 study: (a) distribution of phase change location by treatment arm; (b) distribution of estimated treatment phase (circles) and regrowth phase (triangles) slopes by treatment arm; (c) distribution of estimated treatment phase (circles) and regrowth phase (triangles) curvature parameters by treatment arm.

Table II. Analysis of estimated treatment and regrowth phase slopes (growth rates) as well as time to endpoint (TTE) data.

	Treatment slopes		Regrowth slopes		Time to endpoint effect		
Effect	Estimate	p-Value	Estimate	p-Value	Median	Odds Ratio	p-Value
Control	0.06	<0.001	0.06	<0.001	36.71	*	*
αVEGF	−0.006	0.75	−0.03	0.24	42.37	0.16	0.004
Gem	−0.08	<0.001	−0.01	0.55	79.98	0.04	<0.001
Doc	−0.16	<0.001	0.02	0.22	89	0.01	<0.001
Gem: αVEGF	0.004	0.84	0.03	0.31	68.32	1.05	0.005
Doc: αVEGF	−0.0003	0.98	−0.04	0.17	89	<0.001	1

The first column gives estimated effects. For slopes, variation across treatment arms was analyzed using an ANOVA model ($R^2=0.81$ for treatment phase and $R^2=0.17$ for regrowth phase). The second and fourth columns give estimated effects. The p-values in columns 3 and 5 result from two sided t-tests of significance, with H_0: Effect$=0$. For control, the estimated effect is the average rate for the control arm (used in both treatment and regrowth phases). For other treatments, the estimated effect is a difference from the control rate. The time to endpoint data is analyzed using the Cox regression, yielding estimates of odds ratios of treatment group effects and corresponding p-values.

appear to be substantially lower for all other treatment arms except possibly the αVEGF treatment arm. In order to assess the significance of these differences, we fit an ANOVA-type linear model to the parameter estimates across subjects.

$$b_{gj1}=(1, I[\alpha\text{VEGF}]_{gj}, I[\text{Gem}]_{gj}, I[\text{Doc}]_{\text{gj}}, I[\alpha\text{VEGF}*\text{Gem}]_{\text{gj}}, I[\alpha\text{VEGF}*\text{Doc}]_{gj})^{\text{T}}\beta+\varepsilon_{1gj}. \quad (8)$$

Here b_{gj1} denotes the treatment (first) phase growth rate estimate for the jth subject in the gth group, I stands for an indicator of the relevant treatment, $*$ stands for interaction, β is a vector of treatment effects and ε_{1gj} is i.i.d. the Gaussian error. The estimated effects in Table II show a significant decrease in treatment phase growth rate due to Gem and Doc relative to Control, but no significant decrease due to αVEGF. The effect of Gem during treatment appears to inhibit growth, whereas Doc appears to cause significant reduction in tumor volume during (and some time after) treatment. The interaction effects appear to be small and insignificant. A similar ANOVA of estimated intercepts showed no significant differences across treatments arms.

Regrowth rates (b_2) shown in Figure 4(b) appear to be generally similar to control for all treatment arms except possibly the (Doc+αVEGF) arm. An identical model to (8) was fitted to the estimated regrowth phase rates b_{gj2} (treatment phase rate estimates b_{1j1} were used for the control arm), but Table II shows that no significant main effects or interactions were found. However, since there was specific interest in the effect of combination therapy, two pairwise comparisons were performed. A two sample t-test between (αVEGF+Gem) and Gem alone was not significant (p-value$=0.69$), but (αVEGF+Doc) had a significantly lower regrowth phase rate than Doc alone (mean difference$=0.07$, p-value$=0.03$). This is corroborated in Figure 4(b), where we see that all except one of the estimated regrowth phase rates for the (αVEGF+Doc) are 0, whereas they are mostly positive for the Doc alone treatment arm.

A conventional method for analyzing xenograft experiments is by comparing time to endpoint (TTE), where TTE is the last observed time per subject [3]. Median TTEs show differences between groups, except those for Doc and (αVEGF+Doc), which are identical (Table II). With a multi-group, particularly factorial study design, an obvious extension of log-ranks tests is the Cox regression, assuming proportional hazards across groups. We see that these p-values are generally smaller than in the analysis based on either treatment phase or regrowth phase slopes (Table II). In fact the main effect of αVEGF and αVEGF:gemcitabine is both highly significant in the TTE analysis, whereas they are not significant using slopes. An intriguing result is for the αVEGF:docetaxel interaction, where the TTE p-value is less significant.

Curvature parameters (c_{gj1} and c_{gj2}) for the most treatment arms appear to be close to 0 (Figure 4(c)). The only exceptions are treatment phase curvature parameters for the (αVEGF+Gem) and Gem treatment arms, which appear to be largely negative (growth is slowing). Only the main effect of Gem is significant (Table III). No significant differences in curvature parameter were found between treatment arms during the regrowth phase.

5. Discussion

Apart from facilitating comparisons between treatment arms, piecewise quadratic models can capture tumor growth behavior during both treatment and regrowth phases adequately, as evidenced by high goodness of fit (G_{gjP}) values. The analysis of curvature parameters appears to suggest that the simple exponential model for tumor growth (1) may

Table III. Analysis of estimated treatment and regrowth phase curvature parameters.

	Treatment curvature parameter		Regrowth curvature parameter	
Effect	Estimate	p-Value	Estimate	p-Value
Control	−0.001	0.74	0.0	*
αVEGF	−0.0005	0.90	0.0006	0.40
Gem	−0.02	<0.001	−0.0003	0.64
Doc	−0.0008	0.87	−0.001	0.14
Gem: αVEGF	0.003	0.64	−0.0006	0.54
Doc: αVEGF	0.001	0.78	0.0002	0.83

Variation across treatment arms was analyzed using an ANOVA model ($R^2 = 0.09$ for both treatment and regrowth phase). The second and fourth columns give estimated effects. The p-values in columns 3 and 5 result from two sided t-tests of significance, with H_0: Effect = 0. For control, the estimated effect is the average curvature parameter for the control arm (used in both treatment and regrowth phases). For other treatments, the estimated effect is a difference from control.

be adequate for modeling treatment and regrowth phases for this study. The only exceptions are the arms involving gemcitabine, where the rate of growth appears to be decreasing during the treatment phase. However, since this curvature parameter was typically estimated from only 3 time points, further studies may be necessary to confirm this effect. Nonlinear models have been found to be good fits in other studies [6, 7]. Analysis of curvature parameters may help in the selection of a plausible model in such settings.

The estimation and model fitting procedure presented here has some limitations. At present, our estimation of parameters v_{g1} and λ, which control the influence of the penalty term and hence the positioning of phase change points, is performed in an *ad hoc* manner (Section 2.3). Nevertheless, this method appears to produce consistent estimation of phase change points and slopes in simulation studies (Section 3). Further analysis may help elucidate general conditions under which this occurs. The fitting procedure also assumes that observations within a subject are independent. This may be an unreasonable assumption: residual analysis in the control group (Section 4.2) indicates the presence of mild autocorrelation, which could be due to either (a) correlated measurement errors due to repeat measurements by the same person or (b) the presence of a higher order (than quadratic) trend component. Assuming (a), we can formulate a weighted generalization of the least-squares criterion L using a parsimonious representation of the correlation structure. In simulation studies, we have used an AR(1) process to approximate the observed correlation structure of residual errors.

Estimation of the correlation structure, or parameters thereof, is an important aspect of the analysis of longitudinal data. In smooth models, it is possible to estimate parameters in the correlation structure using criteria such as pseudolikelihood or restricted maximum likelihood (REML) [11]. The optimization of such criteria typically requires iterative algorithms such as iteratively reweighted least squares (IRLS) or Newton–Raphson, which are based on Taylor series expansions of the criterion around the true solution [15]. A critical requirement for the convergence of these algorithms is that the criterion be differentiable in the parameters at the true solution; we note that the piecewise linear model (2) is not differentiable at the phase change points with respect to any of the parameters in the model. We also note that differentiability of the model is a requirement even in non-likelihood-based approaches such as generalized estimating equations [11]. A key advantage of the independence assumption is that it allows the criterion L to be split into sub-criteria L_k which can be minimized separately. This separability is lost in the correlated case, which can lead to non-convergence of the REML-based objective function that we have encountered many times in simulations involving the piecewise model under AR(1) correlation. In this context, it is reassuring to note that when moderate autocorrelation is present, simulation experiments in Section 3 indicate little loss in efficiency for the OLS estimates in the low noise setting relative to a GLS-based approach which incorporates correlation in the fitting criterion. The low noise criterion is satisfied for the data at hand because the quality of fits is extremely good (median R^2 of 0.98). Whether this level of efficiency of estimation for the piecewise parametric model also obtains under correlation structures other than AR(1) remains to be investigated.

A disadvantage of the derived variables analysis approach adopted in this paper is that it does not account for unequal precision of estimated derived variables during the between-subject analysis [11]. For the problem at hand, unequal precision can arise due to the random number of observations per subject caused by volume endpoint censoring, as well as due to adaptive selection of phase change points. We considered using standard errors (SEs) for weighting in the subject wise analysis. However, since SEs are strongly correlated with the number of time points observed, this will tend to bias the average slope downwards, as tumors with lower growth rates will typically have more observed time points. A more sophisticated alternative to this approach is the use of non-linear mixed effect models; however, this approach again involves the use of REML-type criteria which suffer from the same convergence problems for piecewise models

as described in the previous paragraph [16]. An alternative to the piecewise modeling approach is the use of smooth non-parametric ANOVA; this approach uses entire growth curves as units of observation and allows flexibility in shape of the growth curve [17]. However, in situations where the form of the growth curve is well known, as is the case for the control arm, this approach may not yield much extra insight. On the flip side, the variance of non-parametric estimators is typically much higher than those of parametric ones, leading to much less power in testing effects of interest. The piecewise model attempts to incorporate some of the flexibility of the non-parametric approach while still maintaining estimation efficiency by keeping the number of pieces small.

Acknowledgements

The authors thank the assistant editor and referees for their help with improving the paper. Also thanks to Brad Efron, Stanford University, for an insightful discussion on standard errors in derived variables analysis and to Anil Bagri and Leanne Berry, Genentech for design and analysis of the mouse studies. The first author was partially supported in this work by grants from Science Foundation Ireland under Research Frontiers Program grant no. MATF543/2007 and the Maths Initiative/2007.

References

1. Strausberg RL, Simpson AJ, Old LJ, Riggins GJ. Oncogenomics and the development of new cancer therapies. *Nature* 2004; **429**(6990): 469–474.
2. Fujita M, Hayata S, Taguchi T. Relationship of chemotherapy on cancer xenografts in nude mice to clinical response in donor patient. *Journal of Surgical Oncology* 1980; **75**:211–219.
3. Stuschke M, Budach V, Bamberg M, Budach W. Methods for analysis of censored tumor growth delay data. *Radiation Research* 1990; **122**(2):172–180.
4. Skipper H, Schabel F. Quantitative and cytokinetic studies in experimental tumor systems. *Cancer Medicine* (2nd edn). Holland J, Frei E, (eds). Lea and Febiger: Philadelphia, 1982; 636–648.
5. Jagers P. *Branching Processes with Biological Applications*. Wiley: New York, 1975.
6. Demidenko E. *Mixture Models*: *Theory and Applications*. Wiley: New York, 2004.
7. Heitjan D. Generalized Norton Simon models of tumour growth. *Statistics in Medicine* 1991; **10**:1075–1088.
8. Rygaard K, Spang-Thomsen M. Quantitation and Gompertzian analysis of growth. *Breast Cancer Research and Treatment* 1997; **46**:303–312.
9. Voordouw M. Inappropriate application of logarithmic transformations for allometric power functions of morphometric data on acanthocephalan worms. *Journal of Zoological Society of London* 2001; **255**:279–281.
10. Mead R, Curnow R, Hasted A. *Statistical Methods in Agriculture and Experimental Biology* (3rd edn). CRC Press: Boca Raton, 2003.
11. Diggle P, Heagerty P, Liang K, Zeger SL. *Analysis of Longitudinal Data* (2nd edn). Oxford University Press: U.S.A., 2002.
12. Huber PJ. *Robust Statistics*. Wiley: New York, 1981.
13. Serfling RJ. *Approximation Theorems in Mathematical Statistics*. Wiley: New York, 1980.
14. Fuh G, Wu P, Liang W, Ultsch M, Lee C, Moffat B, Wiesmann C. Structure-function studies of two synthetic anti-vascular endothelial growth factor Fabs and comparison with the Avastin Fab. *The Journal of Biological Chemistry* 2006; **281**(10):6625–6631.
15. Giltinan DM, Davidian M. *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall: London, 1995.
16. Lindstrom M, Bates D. Nonlinear mixed effects models for repeated measures data. *Biometrics* 1990; **46**:673–687.
17. Brumback B, Rice J. Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association* 1998; **93**:961–976.