



# Generalized Additive Mixed Modeling of Longitudinal Tumor Growth Reduces Bias and Improves Decision Making in Translational Oncology

William F. Forrest<sup>1</sup>, Bruno Aliche<sup>2</sup>, Oleg Mayba<sup>1</sup>, Magdalena Osinska<sup>3</sup>, Michal Jakubczak<sup>4</sup>, Paweł Piatkowski<sup>5</sup>, Lech Choniawko<sup>6</sup>, Alice Starr<sup>7</sup>, and Stephen E. Gould<sup>2</sup>

## ABSTRACT

Scientists working in translational oncology regularly conduct multigroup studies of mice with serially measured tumors. Longitudinal data collected can feature mid-study dropouts and complex nonlinear temporal response patterns. Parametric statistical models such as ones assuming exponential growth are useful for summarizing tumor volume over ranges for which the growth model holds, with the advantage that the model's parameter estimates can be used to summarize between-group differences in tumor volume growth with statistical measures of uncertainty. However, these same assumed growth models are too rigid to recapitulate patterns observed in many experiments, which in turn diminishes the effectiveness of their parameter estimates as summary statistics. To address this problem, we generalized such models by adopting a nonparametric approach in which group-level response trends for logarithmically scaled tumor volume are estimated as regression splines in a generalized additive mixed model. We also describe a novel summary statistic for group level splines over user-defined,

experimentally relevant time ranges. This statistic reduces to the log-linear growth rate for data well described by exponential growth and also has a sampling distribution across groups that is well approximated by a multivariate Gaussian, thus facilitating downstream analysis. Real-data examples show that this nonparametric approach not only enhances fidelity in describing nonlinear growth scenarios but also improves statistical power to detect interregimen differences when compared with the simple exponential model so that it generalizes the linear mixed effects paradigm for analysis of log-linear growth to nonlinear scenarios in a useful way.

**Significance:** This work generalizes the statistical linear mixed modeling paradigm for summarizing longitudinally measured preclinical tumor volume studies to encompass studies with nonlinear and nonmonotonic group response patterns in a statistically rigorous manner.

## Introduction

Preclinical mouse tumor models serve as a vital bridge between *in vitro* systems and clinical studies because they enable researchers to investigate in-depth hypotheses in a controlled environment while introducing some of the complexity of a living system (1). Such models take many forms ranging in the fidelity to which the system mirrors human pathology (2). A classic and still widely used design for translational studies makes use of subcutaneous or orthotopic xenografts of immortalized cell lines from human tumors in immunodeficient mice (3). Biological artifacts stemming from cell line immortalization and passage in culture can be avoided by directly implanting pieces of human tumors into mice, also known as patient-derived xenografts (PDX;

ref. 4). Studies in cancer immunology require immunocompetency to drive response (5). For such studies, syngeneic grafts of mouse-derived tumor cell lines such as MC38, CT26, or EMT6 are implanted into genetically matched mice (6). Still more complex biology can be investigated through genetically engineered mouse models, which develop tumors stochastically over time, often in the tissues normally associated with the human disease (7). A common feature of these models is that they often yield longitudinal tumor response data (8).

Assessing the impact of novel therapeutics frequently rests on between-group comparisons of longitudinal tumor trends, highlighting the importance of statistical study design and analysis (9). A standardized statistical framework for analysis can underlie objective comparisons across regimens within a study and of interregimen differences across studies. Improvements to the statistical framework can simultaneously enhance decision making (10), reduce the numbers of animals required (11), and support larger translational efforts by enabling summary results from each study to be compared through meta-analysis (12). This article grew out of the need to develop such a framework to analyze experiments from a mouse tumor model-centralized core facility where large numbers of studies are conducted over time. Researchers in this environment need a consistent way to model raw data, evaluate treatment regimens (e.g., drugs with distinct mechanisms or a single drug with different dosing schemes) within a study, and monitor how those regimens' inferred effects compare with historical results. Practically achieving these goals requires a statistical model of tumor growth that conceptually connects the raw longitudinal data to envisioned temporal tumor volume curves and a way to assign summary measures of efficacy to groups based on these curves.

Shortcomings of analyzing single time points from longitudinal data separately have long been understood (13) and have led to the

<sup>1</sup>Department of OMNI Bioinformatics, Genentech, Inc., South San Francisco, California. <sup>2</sup>Department of Translational Oncology, Genentech, Inc., South San Francisco, California. <sup>3</sup>Department of Research Engineering and Software Informatics, Genentech, Inc., South San Francisco, California. <sup>4</sup>Ardigen, S.A., Kraków, Poland. <sup>5</sup>Roche Global IT Solutions Centre: Research and Early Development Support, Roche Pharmaceuticals, Warsaw, Poland. <sup>6</sup>Roche Global IT Solutions Centre: Regions, Diagnostics, and Research Technology Center, Roche Pharmaceuticals, Wrocław, Poland. <sup>7</sup>Insitro, Inc., South San Francisco, California.

W.F. Forrest and B. Aliche contributed equally to this article.

**Corresponding Author:** William F. Forrest, Genentech, Inc., MS 442B, 1 DNA Way, South San Francisco, CA 94080. Phone: 650-225-7261; E-mail: forrest.bill@gene.com

Cancer Res 2020;80:5089–97

**doi:** 10.1158/0008-5472.CAN-20-0342

©2020 American Association for Cancer Research.

development of statistical models of tumor volume that incorporate information across time points to summarize the efficacy of each study regimen. Many such theoretical models of tumor growth exist in the literature. Modeling tumor growth at least over short periods by simple exponential curves has a long history (14) and provides a basis for first-order statistical assessment of treatment effects via, e.g., linear models (15) or linear mixed models (LMM) applied to log-transformed tumor volumes. Exponential growth also serves as a component in more sophisticated schemes using change-point and biexponential methods to assess treatment response (16). A more refined and frequently applied parametric model observes that the growth rate necessarily slows as the tumor itself grows in size. This has inspired a parallel line of work applying Gompertzian growth kinetics to draw insights on tumor biology (17, 18). Other recent proposals include refs. 19, 20, and 21, with a wide-ranging review in ref. 22. Application of such models should also be cognizant of data *dropouts* (23), which describe animals with longitudinal observations stopped prematurely during the study. This can occur for a range of reasons, e.g., an animal is euthanized if the observed tumor grows to be too large or the animal cannot tolerate a treatment; animals can also be taken off study to assess response in one or more biomarkers.

The statistical model in this setting serves two purposes. First, it summarizes and abstracts from the longitudinal data a faithful description of how the tumor volumes change over time for each regimen (essentially a growth curve for each group). These descriptions can be as simple as log-linear growth but are often more complicated with, e.g., periods of tumor regression followed by regrowth, and ideally a model will capture this complexity. Second, the model allows for its by-group descriptions to be distilled into univariate group-level scores for how well or poorly each regimen performs in changing tumor volume growth over time. Comparing these univariate scores allows investigators to quantify the relative efficacy of two or more regimens in a tumor model. For log-linear growth, e.g., such a score can be the log tumor volume growth rate estimated as the slope of a line.

In our experience, a parametric model for tumor growth is chosen for analysis because fitting that model makes it easy to fulfill the second criterion of univariately scoring each group while satisfying the first criterion of describing accurately the tumor growth patterns at least reasonably well. This is a defensible statistical strategy for analysis of any one study in isolation. However, when large numbers of studies each with many different regimens are conducted and analyzed sequentially, it becomes impractical to settle on a common parametric model that plausibly describes all their myriad growth patterns. When a model then notably misfits the data from a treatment regimen, it is reasonable to question whether or how to interpret that model's parametric summary score as a proxy for efficacy.

With this in mind, we refocus on the first purpose of our statistical model and adopt a more flexible, spline-based generalized additive model (GAM; ref. 24) that follows growth trends well even for complicated protocols that induce nonlinear—and sometimes non-monotonic—log tumor volume changes over time. Introduced to integrate splines into classical generalized linear models (GLM), GAMs build on a rich history of research in statistics (25, 26) and provide a statistical framework from which to infer a smooth nonlinear temporal trend for each treatment regimen while retaining many of the advantages of classical GLMs. Although GAMs, like GLMs, can be configured to analyze response data from any exponential family distribution (e.g., Binomial, Poisson, etc.), in this work we restrict attention to responses that are normally distributed about a curve.

Each mouse's tumor volume time series is assumed to be a noisy copy of its group's smooth temporal trend, but potentially with a mouse-specific random affine perturbation (meaning that the trend can randomly “tilt” or vertically “shift” for each mouse). These perturbations extend the GAM to become a generalized additive mixed model (GAMM; ref. 27). GAMMs have been previously applied to longitudinal experiments in translational oncology (28) though without the univariate summary score introduced here. Likelihood-based statistical methods such as LMMs and GAMMs also address the concern posed by dropouts by providing unbiased parameter estimates when applied to longitudinal data with the sorts of dropout patterns typically found in tumor growth studies, where the decision to remove an animal from the study is made based on observed data and a pre-specified protocol (29).

A gap in workflows using GAMMs for such analyses is in distilling each group's estimated spline into a univariate score in a consistent, statistically convenient way. In this article, we propose for such a summary statistic a functional that maps each group's spline fit over a specified time interval to a univariate statistical score through numerical integration. This score reduces to the slope in the case of log-linear tumor growth, which we have found to have intuitive appeal to researchers. Like the slope, its joint summary statistics across a multigroup study can be well-approximated by a multivariate Gaussian distribution with one component per group. This facilitates both intergroup comparisons and interstudy meta-analyses. Because the spline statistic reduces seamlessly to the LMM slope in log-linear scenarios while also quantifying nonlinear ones, studies with a mix of such groups can be analyzed and their treatment regimens scored with a common metric, broadening the applicability of mixed models for tumor growth analysis while effectively retaining the familiar slope summary for linear groups.

In subsequent sections, we describe mouse model tumor experiments employed as examples and the statistical models and summary statistics with which we analyze them. We apply our methods to these experiments in a series of three case studies. Finally, we discuss the results and implications of these analyses as well as advantages and limitations to our proposed approach.

## Materials and Methods

### Mouse oncology models

Analysis methods are demonstrated on data from three tumor growth experiments. The first tracks tumor volumes of mouse-derived allografts from a trichoblastoma cell line, whereas the second and third analyze different drug combinations and dosing schedules in a PDX of glioblastoma multiforme (GBM). In all three studies, each mouse has one subcutaneous tumor that is measured serially by calipers. On each day for which a measurement is taken, the diameter length  $L$  and perpendicular width  $W$  of the tumor are recorded, with volume  $V$  then approximated using the modified ellipsoidal method as  $V = \frac{1}{2}(L \times W^2)$  (30). Tumors are allowed to grow to a volume in an initial range before animals are randomized to treatment groups so as to create closely matched baseline average tumor sizes across regimens. All models were certified to be pathogen-free for *in vivo* use by testing at IDEXX Laboratories, Inc. All *in vivo* studies were conducted in compliance with Genentech's Institutional Animal Care and Use Committee in South San Francisco, CA.

### Murine trichoblastoma allografts

The procedure described in ref. 31 was followed for establishment of a subcutaneous allograft model of murine trichoblastoma with a

loss-of-function mutation in the tumor-suppressor gene *Ptch1*, with human homolog *PTCH1*. Trichoblastoma is a histologic mimic of human basal cell carcinoma (BCC; ref. 32) that serves as a tractable mouse model. The large majority of sporadic BCC clinical cases stem from loss-of-function mutations in *PTCH1* (33), which functions as a natural inhibitor of the *smoothened* (*SMO*) gene. When *PTCH1* is lost, *SMO* becomes dysregulated, leading to abnormal cell growth. A small-molecule inhibitor of *SMO* such as vismodegib (34) can block *SMO* activation. Forty-six female CD1-nude mice (Charles River) with subcutaneous allografts between 195 and 310 mm<sup>3</sup> at baseline were randomized into ten dosing groups, with tumors measured every 3 to 4 days for 21 days. All mice were 10 to 12 weeks of age at study start, with a median body mass of 27.6 g. Each animal was administered an oral dose of vismodegib in MCT vehicle twice daily, at 0, 0.3, 1, 3, 6, 10, 25, 50, 75, or 100 mg/kg.

#### PDX of GBM

Primary human GBM SF7796FL xenografts (35) obtained from the University of California San Francisco were established subcutaneously from trocar-implanted tumor fragments in female NCr nude mice (Taconic). This model was used in two different experiments presented as separate case studies.

#### Palbociclib with and without temozolomide in a PDX of GBM

The reference group in this case study consisted of 7 mice treated orally with 200 mg/kg of palbociclib in lactate buffer (50 mM/L, pH 4.0) once daily for the first 21 days of the study. The active treatment group consisted of 7 mice that received an identical 21-day regimen of palbociclib, but were also administered 22.5 mg/kg of temozolomide in MCT (0.5% methylcellulose, 0.2% Tween-80) orally once daily for the first 5 days of the study. All treatment thus ended for mice in both groups after 21 days. Observations on tumor volume continued for 3 additional weeks beyond this cessation of treatment. All mice were 20 weeks of age at study start, with a median body mass of 29.1 g.

#### Dose escalation of temozolomide on a background of GDC-0084 in a PDX of GBM

In this case study, 20 mice with subcutaneous xenografts between 120 and 245 mm<sup>3</sup> at baseline were randomized into four dosing groups, with tumors measured every 3 to 4 days for 63 days. All mice were 8 weeks of age at study start, with a median body mass of 24.3 g. Mice in the four groups were administered 0, 10, 40, or 80 mg/kg, respectively, of temozolomide in MCT (0.5% methylcellulose, 0.2% Tween-80) orally once daily for the first 5 days of the study. Each animal was also administered an oral dose of GDC-0084 in MCT vehicle once daily at 1 mg/kg for the first 21 days of the study. GDC-0084 is a dual inhibitor of PI3K (36) and mTOR (37). It is under investigation as a therapy for glioblastoma (38), and in August 2019 was granted the proposed international nonproprietary name "paxalisib" by the World Health Organization. All treatment ended for mice in all four groups after 21 days. Observations on tumor volume continued for 6 additional weeks beyond this cessation of treatment, for a total of 9 weeks of longitudinal data.

#### Statistical models

##### Statistical notation

Assume  $I \geq 2$  treatment groups each with  $n_i$  mice ( $i \in 1, \dots, I$ ). Typical values for  $n_i$  are in the range of 4 to 12 mice per group. Each mouse has one tumor, the volume of which is measured at baseline and

then at a number of time points typically over several days. Denote by  $t_{ijl}$  the  $l$ 'th time point for the  $j$ 'th mouse from the  $i$ 'th group, and let  $V_{ijl}$  be the observed volume of a tumor at  $t_{ijl}$ . We apply a variance-stabilizing invertible transformation to the raw tumor volume data and denote by  $y_{ijl}$  the transformed volume at  $t_{ijl}$ . In this work, we log-transform response, setting  $y_{ijl} = \log(V_{ijl} + 1)$ . Other choices (e.g., square roots) can be useful as well; the development below will work with any invertible transformation, though changing the transformation will change the biological interpretation of group-level summary statistics, and we have found logarithms to be suitable in the large majority of cases.

##### Linear and linear mixed effects models

If the log-transformed tumor volumes are linear when graphed against time, a linear model with separate effects by animal and/or treatment group can be both powerful and easy to implement. When each animal's longitudinal profile is cast as a random perturbation of a common group trend, group-level summaries are obtained by LMMs (39). Below we will fit a distinct fixed-effects line for each group with log-scaled tumor volume regressed on time. Each mouse's profile is assumed to have random respective perturbations of its intercept and slope so that the transformed tumor burden  $y_{ijl}$  is modeled as

$$y_{ijl} = (\alpha_i + a_{ij}) + (m_i + b_{ij})t_{ijl} + \epsilon_{ijl}$$

where  $\alpha_i$  and  $m_i$  are the intercept and slope, respectively, for the  $i$ 'th group, whereas  $\{(a_{ij}, b_{ij}), j = 1, \dots, n_i\}$  are the random effects (intercept and slope, respectively) for the tumor profile of the  $j$ 'th animal in the  $i$ 'th group. The random effect terms are assumed to stem from underlying Gaussian distributions, though prediction accuracy of mixed models has been found to be robust to moderate violations of this assumption (40).

##### Generalized additive mixed models to estimate nonlinear growth patterns

Group-level tumor volume profiles are often markedly nonlinear over time, even after logarithmic transformation. To model these nonlinearities, we extend the LMM approach with a GAMM that posits a single flexible growth trend induced in common for each subject in a treatment group, but includes random effects (random intercepts and slopes, in this application) for the profile of each mouse. The log-scaled tumor volume  $y_{ijl}$  is modeled as

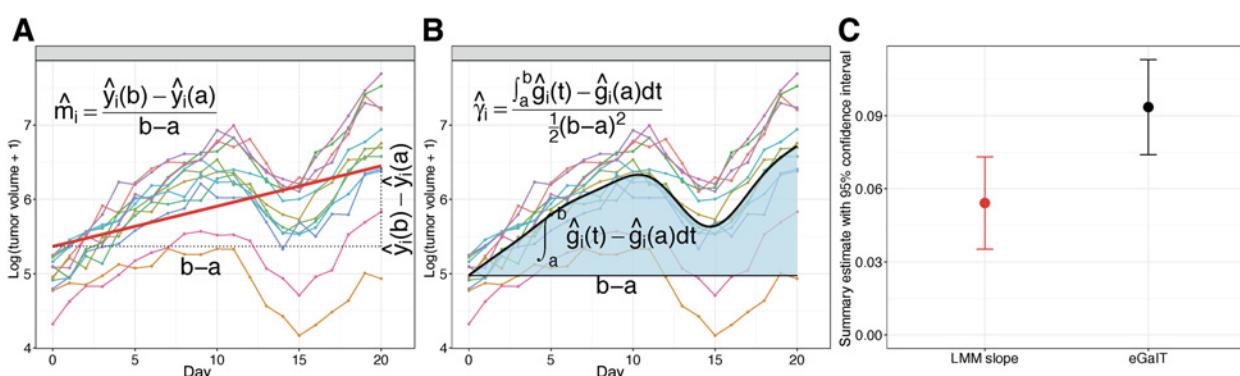
$$y_{ijl} = a_{ij} + b_{ij}t_{ijl} + g(t_{ijl}; \beta_i) + \epsilon_{ijl}$$

where again  $\{(a_{ij}, b_{ij}), j = 1, \dots, n_i\}$  are the random effects for the  $j$ 'th animal's tumor profile, but with the fixed component shared across the  $i$ 'th group now entirely described by a smooth function of time estimated by a spline with parameter vector denoted by  $\beta_i$ . Thin-plate regression splines (41) were used with default parameter settings from the R packages *mgcv* (ref. 24, including, e.g., ten basis vectors, by default) and *gamm4* (42). The resulting GAM spline is denoted below by  $g_i(t)$ ; the spline evaluated at parameter estimates  $\hat{\beta}_i$  is denoted by  $\hat{g}_i(t)$ .

##### Summaries of longitudinal profiles

###### Slopes from a linear mixed model

The LMM growth curve shared by all animals in the  $i$ 'th group is  $E[y_{iji}] = \alpha_i + m_i t_{iji}$ . As efficacy is typically assessed in terms of change from a group's baseline, and because animals are most often randomized so that the baseline average tumor burdens are approximately

**Figure 1.**

**A**, Log-scaled tumor burdens over 20 days with a nonlinear trend and dropouts. The LMM estimate is superimposed in red, and slope estimate  $\hat{m}_i$  is shown. **B**, The same simulated data with GAMM estimate superimposed in black and the eGaIT estimate  $\hat{\gamma}_i$  over 20 days is shown. **C**, A one-way layout to demonstrate that the two estimates  $\hat{m}_i$  and  $\hat{\gamma}_i$  provide distinct but roughly comparable values, with model-based 95% confidence intervals shown.

equal across groups, we take the vector  $(\hat{m}_1, \dots, \hat{m}_I)$  of estimated slopes as summary statistics for regimen efficacy. The linear model can in principle be fit to data that are evidently nonlinear and the slope extracted as a summary, albeit imperfect, of group-level efficacy. A simulated example of such data is shown in Fig. 1A with the linear mixed effects group-level fitted regression line overlaid.

#### Summarizing a GAM fit with the eGaIT statistic

GAMMs are preferable to LMMs for accurately estimating tumor volume growth curves because their greater flexibility can capture nonlinear trends that cannot be accommodated by the simpler linear model. The simplicity of the LMM is advantageous, however, in distilling its by-group fitted lines into regimen summary scores because temporal changes from baseline are captured in the slope estimates and their standard errors. In contrast, a by-group fitted GAM spline from a GAMM lacks an obvious univariate score akin to slope. Across studies, spline bases are determined by data-driven criteria (43) so that the cardinality and meaning of  $\hat{\beta}_i$  can vary. To account for the menagerie of GAM spline models needed to follow patterns across a wide range of tumor volume studies, we propose a regimen summary score that marginalizes across time, intuitively quantifying aspects of a curve shape.

For any time  $t$  in  $[a, b]$ , the growth curve centered to its starting value,  $g_i(t) - g_i(a)$ , is the log fold change gain (or loss) in tumor volume for group  $i$  up through time  $t$ . Adding up the incremental log fold change values across  $[a, b]$  yields the definite integral of  $g_i(t) - g_i(a)$ . We normalize this definite integral by  $\frac{1}{2}(b-a)^2$ . The resulting statistic is the endpoint Gain Integrated in Time ("eGaIT"), denoted by  $\gamma_i$ :

$$\gamma_i \equiv \frac{1}{\frac{1}{2}(b-a)^2} \int_a^b g_i(t) - g_i(a) dt$$

where the estimate  $\hat{\gamma}_i$  is obtained by plugging  $\hat{\beta}_i$  in for  $\beta_i$  in  $g_i(t)$ . Because eGaIT is computed by standard methods in numerical integration, it is ultimately a linear combination of the  $\hat{\beta}_i$  parameter estimates, with standard errors obtained accordingly from the  $\hat{\beta}_i$  values' estimated covariance matrix inferred in the GAMM fitting. It is shown in the Appendix that with the chosen normalizing constant of  $\frac{1}{2}(b-a)^2$ , eGaIT converges seamlessly to the LMM slope as the spline  $g_i(t)$  becomes linear.

The eGaIT statistic is illustrated on the simulated data set in Fig. 1B.

#### Comparing slope and eGaIT

Although the slope  $\hat{m}_i$  assumes underlying log-linear growth, the eGaIT statistic  $\hat{\gamma}_i$  assumes only a smooth response over the study period. Though the eGaIT estimates  $\hat{\gamma}_i$  are scaled so as to align well with their group-matched LMM slope estimates  $\hat{m}_i$  in the log-linear case, they are AUC statistics (44): an alternative interpretation of the eGaIT statistic is as the constant log-linear growth rate that would have been needed to yield the AUC that we actually observed (see Appendix). This interpretation gives some intuition as to when summarizing longitudinal curves by eGaIT is likely to enhance power to discriminate among treatment effects. Curves that over a range of interest are either increasing and concave or decreasing and convex will be more efficiently summarized by the eGaIT statistic, because eGaIT tends to be higher than the slope for increasing but concave growth curves and lower than the slope for decreasing, convex growth curves. Both phenomena are commonly observed for longitudinal tumor data because tumors in the vehicle group frequently show subexponential growth as they become larger (45), whereas tumor size for effective therapies is bounded below so that trajectories for rapidly shrinking tumors tend to flatten out as the study progresses. More generally, eGaIT's formulation typically boosts power for investigating regimens with sustained effects observed from early in the study as opposed to regimens with effects that are transient or manifest late in the study. In this former scenario, eGaIT will be superior to the LMM slope in its power to detect differences between regimens. Examples are presented in the case studies.

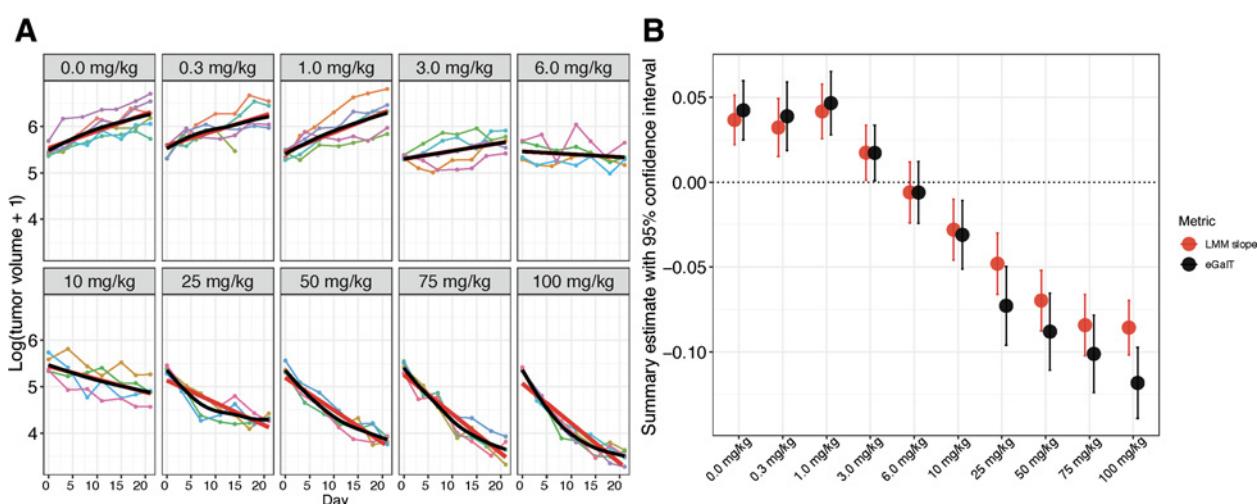
Estimates of the slope  $\hat{m}_i$  and eGaIT  $\hat{\gamma}_i$  with approximate 95% confidence intervals are shown in Fig. 1C.

## Results

We present case studies from three experiments to illustrate how the statistical framework and metrics proposed can be applied in practice to answer questions from translational oncology studies.

#### A dose-escalation study of vismodegib

A ten-group dose-escalation study evaluated vismodegib in 46 CD-1 nude mice with trichoblastoma allografts as described in Materials and Methods. We fit both LMM and GAMM models, calculating slope and eGaIT statistics in order to see where the summary measures match well and where they differ.

**Figure 2.**

**A**, Log-scaled tumor volume data over 21 days from ten groups in a vismodegib dose-escalation study. The red lines are the LMM estimates based on data from days 0 to 21, whereas the black curves are the GAMM estimates over the same time span. **B**, Summary statistics with 95% confidence intervals show the effects on longitudinal tumor volume of vismodegib dose escalation as summarized by the LMM slope and eGaIT statistics, with ten summary measures for each method. The eGaIT statistic provides an extended dynamic range at the higher doses because it captures the signal in the convex, decreasing time-response curve estimates when the dose is at 25 mg/kg or higher. When temporal patterns are linear (10 mg/kg or below), the LMM slope and eGaIT are practically interchangeable summary measures.

**Figure 2A** shows the raw data along with fitted lines and splines from the LMM and GAMM, respectively. We observe linearity at doses that are 10 mg/kg or lower, with slight concavity in the splines at the lowest three doses, and convexity at the four highest doses (25–100 mg/kg). Group-level statistical estimates for the LMM slope and eGaIT with 95% confidence intervals are shown in **Fig. 2B**. Though the two metrics exhibit broad similarity, the eGaIT estimates for the four highest dose groups are all notably lower than the slopes for the same groups, which is a consequence of convexity in the estimated growth curves.

In the six lower-dose groups with essentially log-linear growth, the spline-based statistics and confidence intervals shown in **Fig. 2B** are practically identical to the slope estimates and corresponding confidence intervals from the linear mixed model. This demonstrates empirically that when the data are well summarized by a line, the eGaIT spline summary proposed in this work will effectively reduce to the slope. However, when the data start to show curvature (as in the four highest dose groups), the eGaIT statistic can adapt to and capture the convex response patterns missed by linear estimators, boosting power to detect differences.

#### Temozolomide added to palbociclib in a PDX of GBM

Data from a PDX model of GBM compared treatment over 21 days of palbociclib with and without an initial 5 concurrent days of temozolamide. After day 21 of the study, neither group received any further treatment, but the tumor volumes were recorded for 3 additional weeks, yielding a total of 6 weeks of measurements. In both groups, the tumor growth curves during the 3 weeks of treatment were relatively flat. Over the 3 weeks of posttreatment observation, however, tumors in the group that had received only palbociclib exhibited roughly log-linear growth (**Fig. 3A**, left), whereas the group that in addition received temozolamide on days 0 to 5 showed a marked decline in the tumor volumes (**Fig. 3A**, right).

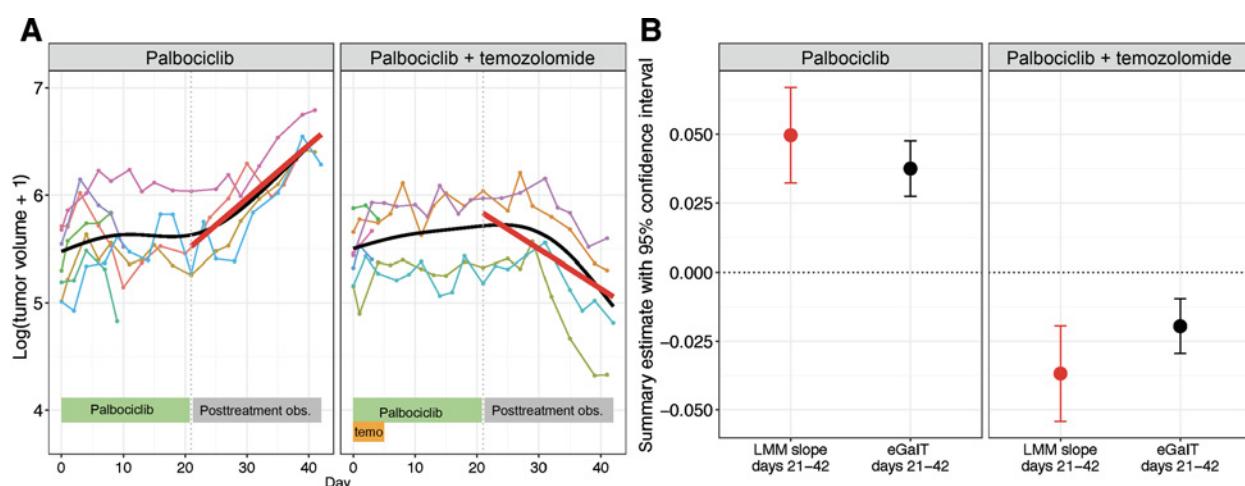
Because palbociclib is a CDK4/6 inhibitor (46) that has been reported to block efficacy of several concurrently dosed chemother-

apeutics (47), it was of interest to compare tumor volume growth between the groups over the posttreatment period (days 21 to 42) in order to characterize the effect of temozolamide dosed on days 0 to 5. One way to do so is to fit a LMM to the data from days 21 to 42 only and report the slopes of the fitted lines on this range. These fitted lines are sketched in red in **Fig. 3A** with slope estimates shown in **Fig. 3B** (in red) and **Table 1**. Another way starts with fitting a GAMM to the entire study interval. Such fitted regression splines in **Fig. 3A** (in black) show mostly static on-treatment tumor volumes for each study group followed by gradually diverging posttreatment growth patterns. Tumor volume growth over the posttreatment period is then estimated from these splines by calculating eGaIT only over days 21 to 42; estimates are shown in **Fig. 3B** (in black) and **Table 1**.

**Table 1** and **Fig. 3B** show that the individual regimen estimates are concordant between the LMM slope and eGaIT, but that the eGaIT estimates are shrunk toward zero and have tighter precision (i.e., narrower confidence intervals). This shrinkage makes eGaIT estimates more realistic summaries of the observed data as it reflects the gradual transitions observable after day 21 from static growth (i.e., a log-fold change of zero) to increasing or decreasing log tumor volume, respectively, for the two regimens.

The effect of initial treatment with temozolamide on tumor growth in the posttreatment period can be summarized either as the difference of the growth rates (slopes) between the two groups in the LMM approach  $\hat{m}_{(\text{palb+temo})} - \hat{m}_{(\text{palb})}$  or as the difference of the eGaIT estimates in the GAMM approach:  $\hat{\gamma}_{(\text{palb+temo})} - \hat{\gamma}_{(\text{palb})}$ . Although eGaIT yields a smaller difference than the LMM slope, its resulting *t*-statistic is greater owing to its tighter precision (**Table 1**).

The improved precision stems from the fact that although eGaIT is estimated over only the posttreatment period, the underlying spline is estimated using data across the full study. This is in contrast to the LMM approach, which fits its model and determines precision of slope estimates from data in the posttreatment period only.

**Figure 3.**

**A**, Data over 42 days of the palbociclib  $\pm$  temozolomide study in a PDX of GBM. The first group was treated with 3 weeks of palbociclib (the green range from days 0 to 21 at the bottom of each plot) followed by 3 weeks of observation (gray range from days 21 to 42). The second group followed an identical regimen, but added 5 days of temozolomide at the study start (orange range in the second plot). The red lines are the LMM estimates based on data from days 21 to 42 only. The black curves are the GAMMM estimates based on data from days 0 to 42. **B**, Summary statistics comparing posttreatment tumor growth in the two groups in the palbociclib  $\pm$  temozolomide study. The red symbols show the LMM slope and 95% CI estimated using the data only from the posttreatment range of days 21–42. The black symbols show the eGaIT statistics estimated from splines fit over days 0–42, but evaluating the difference in endpoint only over days 21–42. The eGaIT estimate matched well with the LMM slope over days 21–42, but was shrunken toward zero and had tighter precision.

#### Temozolomide added to GDC-0084 in a PDX of GBM

A four-group study in a PDX model of GBM compared treatment over 21 days of a common dose of GDC-0084 at 1 mg/kg with an initial 5 concurrent days of four doses of temozolomide, as described in the methods. After day 21 of the study, none of the four groups received any further treatment, but the tumor volumes were recorded for 6 additional weeks, yielding a total of 9 weeks of longitudinal measurements.

**Figure 4A** shows that although the group receiving only GDC-0084 exhibited log-linear growth, the three groups that also received doses of temozolomide (at 10, 40, and 80 mg/kg, respectively) responded with dose-dependent tumor regressions followed by eventual regrowth. It was of interest to quantify each dosing group's degree of regression and regrowth over the 9 weeks of treatment and observation.

The LMM slope is ill-suited to summarizing such marked nonlinearity because the times at the start and end of the measurement range have disproportionate leverage (48) in a linear regression, with the effect that the severe dips and subsequent regrowth measured after treatment with temozolomide are largely irrelevant to the LMM slope. Of the slope estimates with confidence intervals plotted in red in **Fig. 4B**, only that from the 80 mg/kg group has a negative point estimate. None of the LMM slopes is statistically significantly negative,

even though all three nonvehicle groups exhibit tumor shrinkage for most of the 9-week measurement interval.

In contrast, the mid-study dip and regrowth are captured by the GAMMM spline and then incorporated into eGaIT, so that these summary scores (plotted in black in **Fig. 4B**) reflect the evident reductions in volume and dose-dependent delays in regrowth. The eGaIT point estimates over 9 weeks are negative for all three groups dosed with temozolomide, with estimates for the middle and upper doses of temozolomide showing statistically significant reductions.

#### Discussion

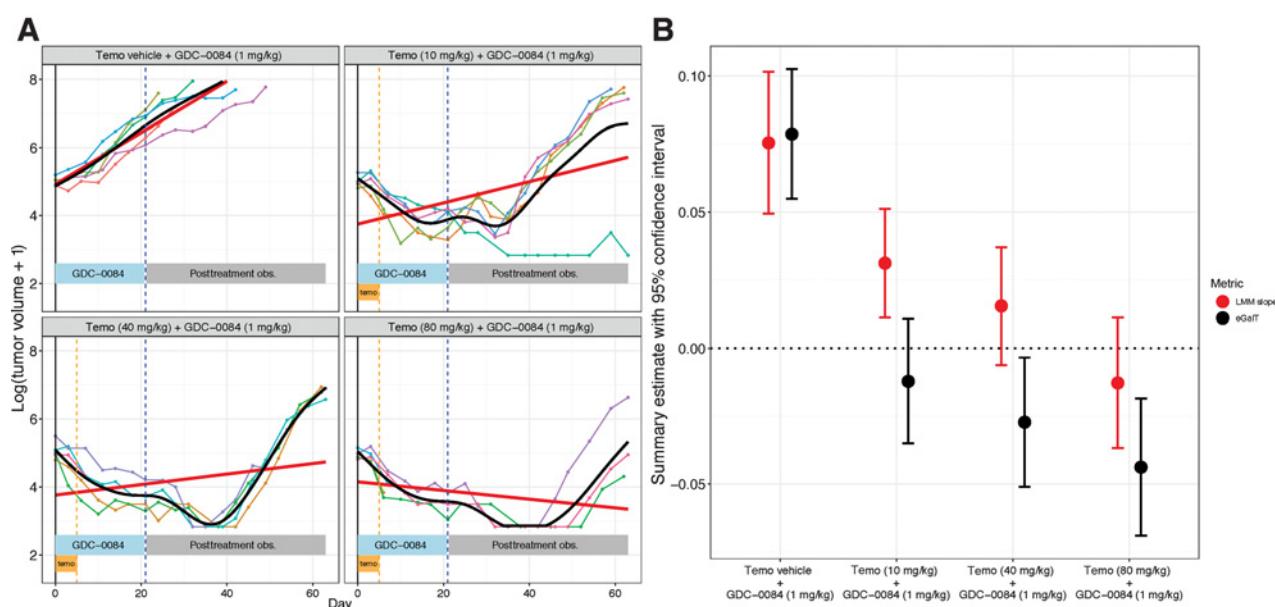
Preclinical tumor studies play a central role in moving novel therapeutics toward the clinic because they provide experimental ground on which biologists and clinicians can test complex hypotheses in controlled settings. Standardized but adaptable analysis of such experiments has been lacking, motivating us to develop a statistical workflow with the two principal goals of estimating the longitudinal tumor volume growth function for each treatment regimen in a flexible, spline-based manner and then summarizing those growth functions across an experimental range of interest so as to yield statistically grounded summary scores amenable to analysis with standard methods and software. Such a biologically motivated summary score that quantifies the qualitative effects evident in nonlinear growth curves while transitioning smoothly to the slope for log-linear growth widens the applicability of longitudinal mixed models by allowing for studies with groups following both linear and nonlinear growth patterns to be analyzed within a common rubric. This workflow benefits researchers by providing them with growth curve estimates that can be superimposed graphically on the raw data to aid in interpreting the study results, and also by giving them standardized and widely applicable statistical summaries with assessments of precision to use in reporting effects and differences across regimens within a study and also across studies (as in, e.g., meta-analyses), and to

**Table 1.** Treatment effects and their difference for LMM slope and eGaIT statistic in palbociclib versus palbociclib + temozolomide.

Metric	Palb (SE)	Palb + temo (SE)	Difference (SE)	t-stat
LMM slope	4.97 (0.78)	-3.68 (0.78)	-8.65 (1.10)	-7.85
eGaIT	3.75 (0.45)	-1.96 (0.44)	-5.71 (0.63)	-9.03

Note: In each case, the SE of the estimate is shown in parentheses. Effect and SE values are all scaled by 100 for legibility, whereas the t-statistic is the ratio of the difference between treatment effects to its SE.

## Generalized Additive Mixed Modeling of Tumor Growth Studies

**Figure 4.**

**A**, Tumor volumes from a PDX of GBM observed over 63 days, with temozolamide dosed over days 0–5 at 0, 10, 40, and 80 mg/kg. In addition, all animals received 1 mg/kg of GDC-0084 over days 0–21. Red lines, LMM estimates; black curves, GAMM estimates. The dotted vertical orange line at day 5 denotes the end of temozolamide dosing. The dotted vertical blue line at day 21 denotes the end of GDC-0084 dosing. **B**, Summary statistics for tumor growth across the four groups receiving temozolamide dose escalation over days 0–5 on a constant background of 1 mg/kg of GDC-0084 over days 0–21. The red symbols show the LMM slope summary and 95% CI estimated over days 0–63. The black symbols show the eGaIT statistics summarizing splines fit over days 0–63.

inform decision making. Although we have presented our method in the context of tumor volume, it can be adapted for analysis of other longitudinal endpoints. Examples from mouse tumor experiments are tolerability proxies such as body weight (49) and imaging-based metrics distinct from tumor volume such as microvascular changes following irradiation (50).

Although slope estimates from LMMs make excellent summaries in cases with clear log-linear growth such as the lower doses of the vismodegib case study, the log-linear approximation is often suboptimal to inadequate when growth of transformed tumor volume is clearly nonlinear as in the two case studies presented with temozolamide combined with different kinase inhibitors. Although some such cases can in principle be handled within the LMM framework by slicing the data into time windows and fitting separate linear models within each (as in, e.g., our case study with palbociclib  $\pm$  temozolamide), that approach not only is more cumbersome but also suffers from several deficiencies. Firstly, mice that drop out in one period are excluded from influencing growth estimates in any subsequent ranges analyzed. Secondly, linear slope estimates from abbreviated time ranges will also suffer from reduced precision, requiring either larger study sizes to begin with or the loss of statistical power. Finally, inferred growth trends estimated separately from adjacent ranges will not in general join smoothly (or even continuously) at the chosen cut points, whereas the by-eye selection of cut points to identify linear response ranges can itself introduce overly optimistic or biased slope estimates.

In contrast, the application of GAMM-based regression splines as described here allows the growth curves to be estimated by a set of continuous functions based on all the data available. Hypotheses about how regimens compare over ranges can then be summarized and reported with model-based measures of uncertainty. Our proposed

eGaIT statistic provides biologically motivated summaries of fitted splines that coincide with the intuitively appealing and widely used slope in the log-linear case but generalize usefully in nonlinear scenarios such as the case study of concurrent dosing with temozolamide and GDC-0084. Although the resulting modeling scheme is not optimal for all designs, it has proven to be at least adequate for the vast majority of studies and provides a major simplification by allowing us to automate robust and flexible modeling and summarization of growth functions that can be applied by oncology researchers for standard analyses while consulting a statistician only as needed.

There are several limitations to our approach with varying levels of impact. One is the assumption that each animal's tumor growth curve within a group follows a common pattern (with random effects and noise). This assumption can fail in studies with extensive within-regimen heterogeneity of response, observed in particular with immune-based therapies. In this setting, a subset of the tumors may have partial or even complete responses while others have growth indistinguishable from a control group. In such a case, estimating the *average* response (linear or nonlinear) is ineffective in summarizing volume changes because few if any of the tumor growth profiles will be well described by that average. One alternative is to enumerate responders within each group and employ methods from categorical data analysis (51). Our software includes functions to facilitate such analyses, with examples in the package documentation. Another limitation of the methods presented here is the need for a study to have enough data to fit a spline. Though we have found the underlying R packages to be impressively resilient, studies with few time points (e.g., three distinct days) can result in nonidentifiability or numerical instability that precludes fitting splines, though trends in such a study are still nonlinear. In such cases, piecewise

linear mixed models (24) have often sufficed as an alternative for curve fitting; these are also implemented in our software but are outside the scope of the current discussion. Finally, common spline smoothing terms and variance components for random effects are estimated and applied across the groups. This is most often a reasonable assumption, and the choice was a deliberate trade-off, as keeping the fitted splines within a common statistical model greatly eases incorporating measures of uncertainty in the numerical integrations underlying our summary statistics.

We note that fitting a spline with a GAMM and then summarizing that spline by eGaIT are distinct and modular steps, so that alternative spline summary statistics can be defined and encoded to capture different qualities of a growth trajectory in cases when eGaIT may not suffice. One such alternative spline summary is the *endpoint Difference Over Time* (eDOT), which is the difference in spline values at the study interval endpoints divided by the interval length. This more directly generalizes the concept of linear slope to a nonlinear response, though it relies only on the spline's value at the start and end of the study, ignoring its interim path. Our R package available at <https://github.com/wfforrest/maeve> enables streamlined fitting of GAMMs to multigroup studies and by-group summarization of the fitted splines to either of eGaIT or eDOT, with further details on both available in package documentation. Included with the R package documentation are example workflows on real data and a characterization study that includes simulations in which confidence intervals for eGaIT and eDOT exhibit nearly nominal coverage for group sizes similar to those of the case studies shown here.

## Appendix: The Normalizing Constant for eGaIT

The choice of  $\frac{1}{2}(b-a)^2$  as a constant of normalization for  $\gamma$  is motivated by examining a Taylor expansion of  $g$  about  $t = a$ . Dropping the subscript  $i$  and letting  $g^{(k)}(x)$  denote the  $k$ 'th derivative of  $g$ , we have

$$g(t) - g(a) \approx g^{(1)}(a)(t-a) + \sum_{k \geq 2} \frac{g^{(k)}(a)}{k!} (t-a)^k$$

so that the baseline-normalized AUC over  $t \in [a, b]$  is

$$\int_a^b g(t) - g(a) dt \approx g^{(1)}(a) \int_a^b (t-a) dt + \sum_{k \geq 2} \frac{g^{(k)}(a)}{(k+1)!} (b-a)^{k+1}$$

The first definite integral on the right evaluates to  $\frac{1}{2}(b-a)^2$ . By choosing  $\frac{1}{2}(b-a)^2$  as the normalizing constant, the statistic  $\gamma$

## References

- Gengenbacher N, Singhal M, Augustin HG. Preclinical mouse solid tumour models: status quo, challenges and perspectives. *Nat Rev Cancer* 2017;17:751.
- Gould SE, Junttila MR, de Sauvage FJ. Translational value of mouse models in oncology drug development. *Nat Med* 2015;21:431–9.
- Pantelouris E. Absence of thymus in a mouse mutant. *Nature* 1968;217:370.
- Byrne AT, Alférez DG, Amant F, Annibali D, Arribas J, Biankin AV, et al. Interrogating open issues in cancer precision medicine with patient-derived xenografts. *Nat Rev Cancer* 2017;17:254.
- Selby MJ, Engelhardt JJ, Johnston RJ, Lu L-S, Han M, Thudium K, et al. Preclinical development of ipilimumab and nivolumab combination immuno-therapy: mouse tumor models, in vitro functional studies, and cynomolgus macaque toxicology. *PLoS One* 2016;11:e0161779.
- Yang Y, Yang HH, Hu Y, Watson PH, Liu H, Geiger TR, et al. Immunocompetent mouse allograft models for development of therapies to target breast cancer metastasis. *Oncotarget* 2017;8:30621.
- Kersten K, de Visser KE, van Miltenburg MH, Jonkers J. Genetically engineered mouse models in oncology research and cancer medicine. *EMBO Mol Med* 2017;9:137–53.
- Jensen MM, Jørgensen JT, Binderup T, Kjær A. Tumor volume in subcutaneous mouse xenografts measured by microCT is more accurate and reproducible than determined by 18F-fdg-microPET or external caliper. *BMC Med Imaging* 2008;8:16.

## Generalized Additive Mixed Modeling of Tumor Growth Studies

9. Laajala TD, Jumppanen M, Huhtaniemi R, Fey V, Kaur A, Knuutila M, et al. Optimized design and analysis of preclinical intervention studies in vivo. *Sci Rep* 2016;6:30723.
10. Begley CG, Ellis LM. Drug development: raise standards for preclinical cancer research. *Nature* 2012;483:531–3.
11. Burden N, Chapman K, Sewell F, Robinson V. Pioneering better science through the 3Rs: an introduction to the national centre for the replacement, refinement, and reduction of animals in research (nc3rs). *J Am Assoc Lab Anim Sci* 2015;54:198–208.
12. Jue TR, Sena ES, Macleod MR, McDonald KL, Hirst TC. A systematic review and meta-analysis of topoisomerase inhibition in pre-clinical glioma models. *Oncotarget* 2018;9:11387.
13. Heitjan DF, Manni A, Santen RJ. Statistical analysis of in vivo tumor growth experiments. *Cancer Res* 1993;53:6042–50.
14. Skipper HE. Experimental evaluation of potential anticancer agents xiii, on the criteria and kinetics associated with "curability" of experimental leukemia. *Cancer Chemother Rep* 1964;35:3–111.
15. Wu J. Assessment of antitumor activity for tumor xenograft studies using exponential growth models. *J Biopharm Stat* 2011;21:472–83.
16. Zhao L, Morgan MA, Parsels LA, Maybaum J, Lawrence TS, Normolle D. Bayesian hierarchical changepoint methods in modeling the tumor growth profiles in xenograft experiments. *Clin Cancer Res* 2011;17:1057–64.
17. Norton L. A gompertzian model of human breast cancer growth. *Cancer Res* 1988;48:7067–71.
18. Dhruba SR, Rahman A, Rahman R, Ghosh S, Pal R. Recursive model for dose-time responses in pharmacological studies. *BMC Bioinformatics* 2019;20:317.
19. Hartung N, Mollard S, Barbolosi D, Benabdallah A, Chapuisat G, Henry G, et al. Mathematical modeling of tumor growth and metastatic spreading: Validation in tumor-bearing mice. *Cancer Res* 2014;74:6397–407.
20. Murphy H, Jaafari H, Dobrovolsky HM. Differences in predictions of ODE models of tumor growth: a cautionary example. *BMC Cancer* 2016;16:163.
21. Novick SJ, Sachsenmeier K, Leow CC, Roskos L, Yang H. A novel Bayesian method for efficacy assessment in animal oncology studies. *Stat Biopharm Res* 2018;10.
22. Benzekry S, Lamont C, Beheshti A, Tracz A, Ebos JM, Hlatky L, et al. Classical mathematical models for description and prediction of experimental tumor growth. *PLoS Comput Biol* 2014;10:e1003800.
23. Martin EC, Aarons L, Yates JW. Accounting for dropout in xenografted tumour efficacy studies: integrated endpoint analysis, reduced bias and better use of animals. *Cancer Chemother Pharmacol* 2016;78:131–41.
24. Wood S. Generalized additive models: an introduction with R. 2nd ed. Boca Raton, Florida: Chapman Hall/CRC; 2017.
25. Wahba G. A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *Ann Stat* 1985;13:78–402.
26. Friedman JH, Silverman BW. Flexible parsimonious smoothing and additive modeling. *Technometrics* 1989;31:3–21.
27. Lin X, Zhang D. Inference in generalized additive mixed models by using smoothing splines. *J R Stat Society Series B Stat Methodol* 1999;61:381–400.
28. Liang H. Modeling antitumor activity in xenograft tumor treatment. *Biom J* 2005;47:358–68.
29. Fitzmaurice GM, Laird NM, Ware JH. Applied longitudinal analysis. Hoboken, New Jersey: John Wiley & Sons; 2012.
30. Kersmans V, Cornelissen B, Allen PD, Beech JS, Smart SC. Subcutaneous tumor volume measurement in the awake, manually restrained mouse using MRI. *J Magn Reson Imaging* 2013;37:1499–504.
31. Wang GY, So P-L, Wang L, Libove E, Wang J, Epstein EH Jr. Establishment of murine basal cell carcinoma allografts: A potential model for preclinical drug testing and for molecular analysis. *J Invest Dermatol* 2011;131:2298–305.
32. Stanoszek LM, Wang GY, Harms PW. Histologic mimics of basal cell carcinoma. *Arch Pathol Lab Med* 2017;141:1490–502.
33. Epstein EH. Basal cell carcinomas: attack of the hedgehog. *Nat Rev Cancer* 2008;8:743.
34. Sekulic A, Migden MR, Oro AE, Dirix L, Lewis KD, Hainsworth JD, et al. Efficacy and safety of vismodegib in advanced basal-cell carcinoma. *N Engl J Med* 2012;366:2171–9.
35. Hadaczek P, Ozawa T, Sorceanu L, Yoshida Y, Matlaf L, Singer E, et al. Cidofovir: a novel antitumor agent for glioblastoma. *Clin Cancer Res* 2013;19:6473–83.
36. Fruman DA, Chiu H, Hopkins BD, Bagrodia S, Cantley LC, Abraham RT. The PI3K pathway in human disease. *Cell* 2017;170:605–35.
37. Xie J, Wang X, Proud CG. MTOR inhibitors in cancer therapy. *F1000Research* 2016;5.
38. Ippen FM, Alvarez-Breckenridge CA, Kuter BM, Fink AL, Bihun IV, Lastrapes M, et al. The dual PI3K/mTOR pathway inhibitor GDC-0084 achieves antitumor activity in PIK3CA-mutant breast cancer brain metastases. *Clin Cancer Res* 2019;25:3374–83.
39. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982;38:963–74.
40. McCulloch CE, Neuhaus JM. Prediction of random effects in linear and generalized linear models under model misspecification. *Biometrics* 2011;67:270–9.
41. Wood SN. Thin plate regression splines. *J R Stat Soc Series B Stat Methodol* 2003;65:95–114.
42. Wood S, Scheipl F. Gamm4: generalized additive mixed models using 'mgcv' and 'lme4' [Internet]. 2016. Available from: <https://CRAN.R-project.org/package=gamm4>
43. Wood SN, Pya N, Safken B. Smoothing parameter and model selection for general smooth models. *J Am Stat Assoc* 2016;11:1–45.
44. Paulson JN, Talukder H, Bravo HC. Longitudinal differential abundance analysis of microbial marker-gene surveys using smoothing splines. *BioRxiv* 2017;099457.
45. Norton L. Theoretical concepts and the emerging role of taxanes in adjuvant therapy. *Oncologist* 2001;6:30–5.
46. Liu M, Liu H, Chen J. Mechanisms of the CDK4/6 inhibitor palbociclib (pd 0332991) and its future application in cancer treatment. *Oncol Rep* 2018;39:901–11.
47. McClendon AK, Dean JL, Rivadeneira DB, Yu JE, Reed CA, Gao E, et al. CDK4/6 inhibition antagonizes the cytotoxic response to anthracycline therapy. *Cell Cycle* 2012;11:2747–55.
48. Chatterjee S, Hadi AS. Influential observations, high leverage points, and outliers in linear regression. *Stat Sci* 1986;1:379–93.
49. Aston WJ, Hope DE, Nowak AK, Robinson BW, Lake RA, Lesterhuis WJ. A systematic investigation of the maximum tolerated dose of cytotoxic chemotherapy with and without supportive care in mice. *BMC Cancer* 2017;17:684.
50. Demidov V, Maeda A, Sugita M, Madge V, Sadanand S, Flueraru C, et al. Preclinical longitudinal imaging of tumor microvascular radiobiological response with functional optical coherence tomography. *Sci Rep* 2018;8:38.
51. Santner TJ, Duffy DE. The statistical analysis of discrete data. New York: Springer-Verlag; 1989. p.367.

# Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

## Generalized Additive Mixed Modeling of Longitudinal Tumor Growth Reduces Bias and Improves Decision Making in Translational Oncology

William F. Forrest, Bruno Aliche, Oleg Mayba, et al.

*Cancer Res* 2020;80:5089-5097. Published OnlineFirst September 25, 2020.

**Updated version** Access the most recent version of this article at:  
doi:[10.1158/0008-5472.CAN-20-0342](https://doi.org/10.1158/0008-5472.CAN-20-0342)

**Cited articles** This article cites 42 articles, 8 of which you can access for free at:  
<http://cancerres.aacrjournals.org/content/80/22/5089.full#ref-list-1>

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, use this link  
<http://cancerres.aacrjournals.org/content/80/22/5089>.  
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.