

Characterization of two summary statistics for fitted splines from generalized additive mixed models.

William F. Forrest*, Bruno Aliche*, Oleg Mayba, Magdalena Osinska,
Michał Jakubczak, Paweł Piatkowski, Lech Choniawko, Alice Starr,
Stephen E. Gould

2020-08-25

Abstract

Some theoretical background on a generalized additive mixed model applied to longitudinal data is presented, and two summary statistics for the fitted regression splines from this model described and characterized. A simulation study is performed in which underlying log tumor volume growth is linear in a control group but nonlinear in a treatment group, with data subject to longitudinal dropouts or not. Three summary estimates of the resulting longitudinal data are computed in each case: the fixed effect slope from a linear mixed model and two different spline summary statistics from a generalized additive mixed model. The simulation demonstrates that when the underlying growth trajectory is nonlinear, the sampling distribution of the slope can change noticeably when the data contain dropouts. In contrast, the spline summary estimates both have sampling distributions that are relatively robust to effects of dropouts. Spline summaries also maintain their confidence interval coverage rates at nearly the nominal level in the presence of nonlinearity and / or dropouts, while the slope estimates do not.

Contents

1	Introduction	2
2	Details of the implemented spline model	2
3	Details of spline summary measures	3
3.1	endpoint Gain Integrated in Time (eGaIT)	3
3.2	endpoint Difference Over Time (eDOT)	3
3.3	numerical integration for inference on spline summaries	4
4	A simulation study to characterize group-level curve summary statistics	5
4.1	Growth model	5
4.2	True values of summary metrics	6
4.3	Effects on the estimators of non-uniform measurement times.	7
5	Results	8
5.1	Point estimates of LMM slope, eGaIT, and eDOT	8
5.2	Confidence interval coverage	12
6	Discussion	13
7	References	14
8	Appendix	15
8.1	Appendix: Normalizing constants for spline summary statistics	15

1 Introduction

This document first explains some technical aspects of the regression spline mixed model applied to tumor growth studies, then explores through simulation the sampling distributions and performances of the summary estimators for both linear and nonlinear growth patterns. We characterize the estimators when data are complete, then also when some time points are excluded entirely or when some mice drop out of the study early. Fitting of the linear and generalized additive mixed models and calculation of slopes and summary statistics are orchestrated through `maeve`, an R package we have written and made available at <https://github.com/wfforrest/maeve>.

We will see that the spline summary statistics provide nearly unbiased estimates of their true underlying parameter values even when the true pattern is nonlinear and the data contain dropouts. Model-based 95% confidence intervals for these parameters consistently cover the true parameter values at slightly less than the nominal rate (e.g., 92% coverage for 5 animals per group, in our setting), so that while imperfect, they seem a reasonable basis for decision-making. In contrast, the linear mixed model coverage is reliable when the transformed growth is linear without dropouts, but erratic when used to approximate a nonlinear scenario, with sampling distributions for the slope changing markedly depending on the spacing of observation times and patterns of dropouts in the data. Since real studies usually have dropouts within groups that may themselves follow nonlinear growth trends, the generalized additive mixed model can avoid vagaries of effects within and across studies due, e.g., to differences in scheduled measurement times or fractions of dropouts.

2 Details of the implemented spline model

Spline-based modeling has a rich history in statistics (1), (2), (3), (4) with applications in a diverse range of scientific problems (5) (6) (7). In modeling tumor growth data, we decided to build on work done in generalized additive mixed models (GAMMs) since they provide a natural extension of linear mixed models (LMMs) that have been applied successfully to time course data with mixed effects by subject when trends are linear. A frequent technical issue in applying spline-based methods is the placement of spline “knots” across the interval of interest. This was a point of concern in our application as we sought to minimize – and ideally avoid altogether – knot-setting so as to simplify the functions and avoid arbitrariness in downstream results based on numbers and locations of knots.

To this end, we built our approach using the `mgcv` package (8) in R, which was found to compare favorably to a number of alternatives (9). Among several choices, `mgcv` has well-developed functionality for implementing a *thin plate regression spline* (TPRS) approach, which circumvents the question of knot placement. Intuitively, TPRSs start with a very large collection of polynomial vectors, where each vector is centered at one of the independent “x” predictors associated with a data point, but then extract a low-rank eigen-decomposition from this initial panoply of vectors and take this eigen set as the spline basis vectors in a regression (10), with smoothing accomplished by either of generalized cross validation or by restricted maximum likelihood (which treats the spline coefficients as random effects). TPRS models are described thoroughly in (11), and applied in a detailed case study in (12).

The package `gamm4` (13) provides a simplified interface, similar to that for the `lmer` package (14), for fitting generalized additive mixed models with the underlying TPRS splines determined within the `mgcv` package. Rather than specifying numbers and locations of “knots”, a thin plate regression requires a number of basis vectors. This number can be set manually within the package, but the default settings usually work well in our experience unless the number of distinct time points is quite low (e.g., below 4), in which case identifiability is an issue and users may need to consider implemented alternatives such as piecewise linear mixed models, described in a vignette for `maeve` but beyond the scope of the work presented here.

Longitudinal tumor studies typically feature at least some dropouts and often include at least one group with nonlinear growth over the temporal range of interest, so that summary methods robust to nonlinearity and dropouts are desirable. In this characterization study we present a simple two-group simulation in order to demonstrate in this setting that the summary statistics for the GAMM perform as well or better than LMM slope estimates in the presence of nonlinearity and / or dropouts.

3 Details of spline summary measures

For the i 'th group in a study, the average transformed tumor volume trajectory follows a growth function that is assumed to be well approximated by a GAM spline $g(t; \beta_i)$, denoted below by $g_i(t)$, where β_i is a p -vector of spline coefficients. The spline evaluated at parameter estimates $\hat{\beta}_i$ is denoted by $\hat{g}_i(t)$. In this section, we revisit the eGait statistic of (15) then introduce a second spline summary statistic called eDOT, also implemented in our software package. In an appendix, we also provide heuristic arguments for why the two spline summaries (eGait and eDOT) will both converge to the LMM slope as log-scale tumor volume growth patterns become linear.

3.1 endpoint Gain Integrated in Time (eGait)

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 γ_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 spline coefficient estimates $\hat{\beta}_i$ in for β_i in $g_i(t)$.

3.2 endpoint Difference Over Time (eDOT)

Instead of the eGait summary statistic, an alternate summary of efficacy for a spline $g_i(t)$ is the time-averaged tumor growth *rate* over $[a, b]$, *i.e.*, the ratio of the time-integrated growth rate to the study length. By the fundamental theorem of calculus, this is the difference $g_i(b) - g_i(a)$ in endpoint values divided by the study length. We refer to this time-normalized difference as the *endpoint Difference Over Time* (eDOT), denoted by δ_i :

$$\delta_i \equiv \frac{1}{b-a} \int_a^b g'_i(t) dt = \frac{g_i(b) - g_i(a)}{b-a}$$

where the eDOT estimate $\hat{\delta}_i$ is obtained by plugging $\hat{\beta}_i$ in for β_i in $g_i(t)$. The estimate will track closely with the slope of a fitted line in the case that growth is roughly linear. The vector of estimates across the I groups is denoted $\hat{\Delta}_{1 \times I} \equiv (\hat{\delta}_1, \dots, \hat{\delta}_I)$ and is approximately multivariate Gaussian.

The eDOT and eGait statistics embody different views of how to summarize the observed tumor volume changes. Put succinctly, eDOT depends only on where average tumor volume begins and ends, while eGait depends also on the growth *path* the tumors took between that start and end. In the case of log-linear growth, the LMM slope, eDOT, and eGait yield practically identical treatment summaries, providing a base point from which to regard eDOT and eGait as distinct ways to generalize log-linear growth by slopes. In an appendix, we include mathematical rationales for the respective normalizing constants of eGait and eDOT

so that they converge smoothly to the LMM slope in studies for which the log tumor volume growth rate is linear.

In our experience eDOT often yields a value quite close to the LMM slope estimate, even with some nonlinearities in the longitudinal responses. This is unsurprising since it generalizes to splines the “rise over run” mnemonic from elementary algebra for the slope of a line, which made eDOT a natural initial choice for summarizing spline trajectories. In practice, though, we found that its indifference to the growth path taken over a study led eDOT to inherit many of the suboptimal properties of the LMM slope as a summary statistic, so that eGait became the preferred summary in our setting.

3.3 numerical integration for inference on spline summaries

Fitting the GAMM yields, for each group, a p -vector $\hat{\beta}$ of basis spline coefficient fixed effect estimates and their corresponding $p \times p$ covariance estimate $\hat{\Sigma}$ (which is common across groups). We sketch here how to obtain eGait or eDOT estimates and standard errors for a group’s fitted spline over a specified range.

1. Pick a series of equally-spaced x -values $\{x_1^*, x_2^*, \dots, x_k^*\}$. The “*” superscripts are included to emphasize that these do not need to match any of the independent x values observed in the real data. The value of k is the grid size in a numerical integration below. This can be set in the software, but we have found a default value of 25 to provide good performance.
2. Evaluate x_i^* (for $i \in 1, \dots, k$) at each of the p thin plate spline basis vectors and arrange the values into a $1 \times p$ row vector. Stack the K row vectors to assemble a $k \times p$ design matrix X^* .
3. Post-multiplying the design matrix X^* by the estimated spline effects $\hat{\beta}$ defines a k -vector of spline predicted values $Y^* \equiv X^* \hat{\beta}$.
4. The area under the curve (AUC) can be estimated numerically by, e.g., Newton-Cotes quadrature (we use Simpson’s Rule). Let $\{w_1, w_2, \dots, w_k\}$ be the real-valued weights derived from the $\{x_1^*, x_2^*, \dots, x_k^*\}$ values, and denote by w the k -vector of these quadrature weights. Then, an estimate of the AUC is $w^T Y^* = w^T X^* \hat{\beta}$.
5. Finally, define a contrast p -vector $c = (X^*)^T w$ so that $c^T \equiv w^T X^*$. The AUC can now be written as $c^T \hat{\beta}$, the inner product of c with the p -vector of spline basis vector fixed effect estimates $\hat{\beta}$. Note that c depends on the spline basis vectors in X^* and on the grid $\{x_1^*, x_2^*, \dots, x_k^*\}$, but not on the responses.
6. There are a number of details not exhaustively described here. Two key special cases, however, are:
 - When subtracting out the curve’s baseline value from the AUC, define a second $k \times p$ design matrix X_1^* in which each row is just the p -vector of spline basis functions evaluated at the left-most grid point x_1^* , then work with the $k \times p$ matrix $X^* - X_1^*$ as the design matrix in the steps above.
 - A similar trick facilitates finding the integrated first derivative while retaining the structure to incorporate covariance. Define a second $k \times p$ design matrix $X_{\Delta x}^*$ with entries found by evaluating right-shifted grid values $x_i^* + \Delta x$ in each spline basis vector for a small value of Δx , then use $\frac{1}{\Delta x}(X_{\Delta x}^* - X^*)$ as the design matrix in the steps above. (The code implementation makes modest adjustments to gain some accuracy and deal with interval edge cases, but this is the basic idea).
7. Given $\hat{\beta}$, $\hat{\Sigma}$, and the contrast p -vector c , standard results from linear regression are used to estimate numerical integrals and their standard errors. E.g., in the simplest case, AUC can be estimated by $c^T \hat{\beta}$ with a standard error estimate of $(c^T \hat{\Sigma} c)^{\frac{1}{2}}$. More complicated comparisons can be obtained by assembling and combining particular contrast matrices. We leverage capabilities from the R package `multcomp` to implement Dunnett contrasts (many active treatments compared to a common control), Tukey contrasts (all pairwise comparisons), and some customized contrast, with examples in the package vignette.

4 A simulation study to characterize group-level curve summary statistics

We simulate log-scale longitudinal data from both linear and nonlinear growth curves over an interval so that true values of the curve summaries (LMM slope, eGaIT, eDOT) can be computed analytically and thus are known as gold-standard benchmarks. We then apply a number of filtering and dropout criteria to assess the effects that uneven spacing of observation times and dropouts have on each estimator, and repeat the simulation for two different per-group sample sizes in order to assess how well the asymptotic approximations used in inference hold up in (simulated) practice.

4.1 Growth model

We simulate tumor sizes measured from mice in each of two groups. The control group tumors grow log-linearly from baseline over twenty days, with a growth curve of log-scale tumor volume denoted $g_0(t)$:

$$g_0(t) = \frac{1}{10}t + 5$$

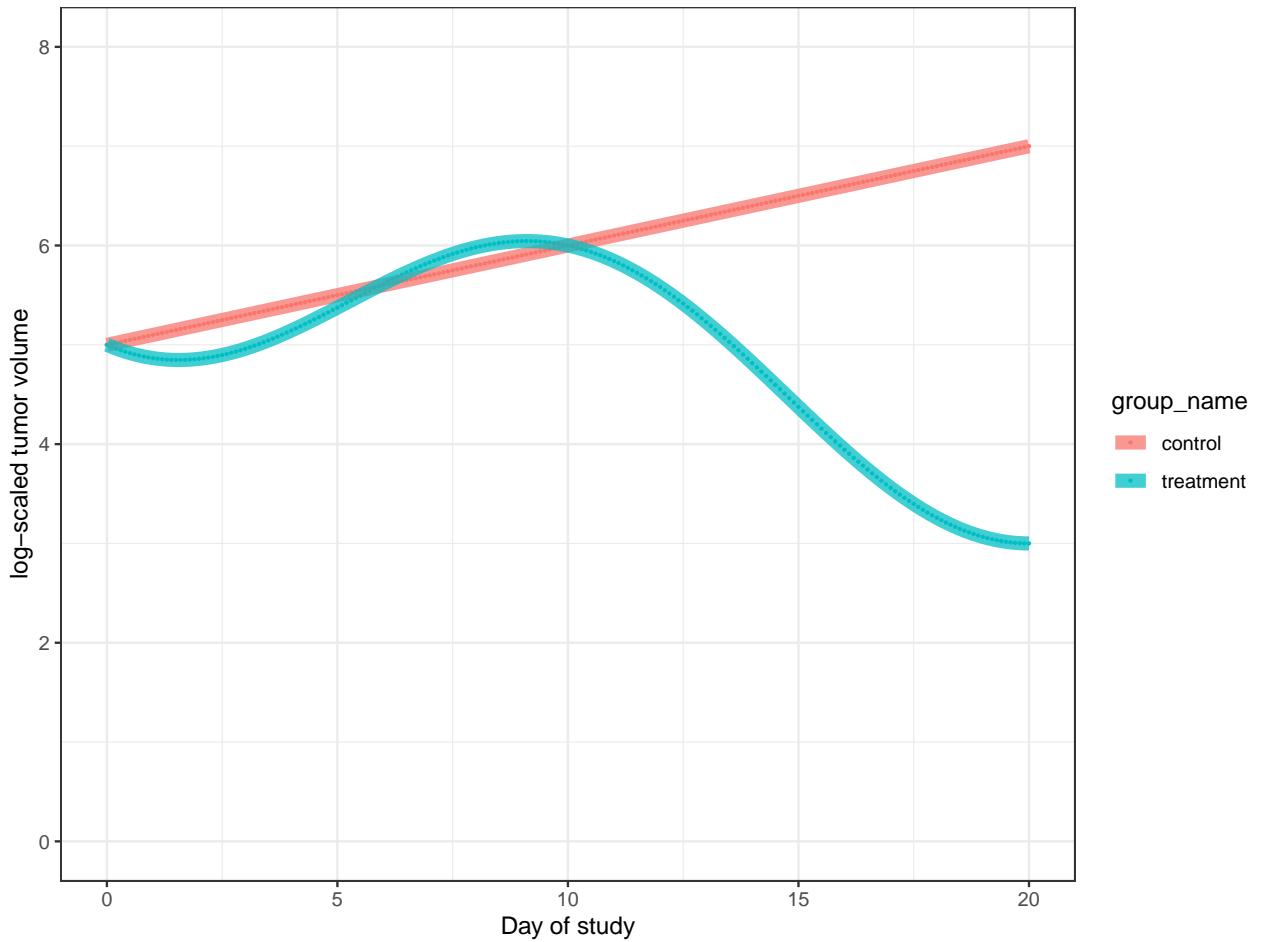
so that the expected log size on day $t = 0$ is 5, and on day $t = 20$ is 7.

The treatment group tumors grow nonlinearly on the logarithmic scale from baseline over twenty days, with a growth curve $g_1(t)$ defined by an analytically tractable function: a convex quadratic superposed with a scaled and shifted cosine curve that starts and ends our twenty-day simulation at a height of zero but grows and recedes in the interim:

$$g_1(t) = \frac{1}{200}t^2 + \frac{-1}{5}t + 5 + \frac{5}{4} \left[1 - \cos \left(2\pi \frac{t}{20} \right) \right]$$

so that its expected log size on day $t = 0$ is 5, and on day $t = 20$ is 3. The true curves are shown in Figure S.1.

Figure S.1



4.2 True values of summary metrics

4.2.1 LMM slope

The true underlying slope for the control curve $g_0(t)$ is $m_0 = .10$.

For the nonlinear treatment curve $g_1(t)$, the underlying slope m_1 is not defined as a constant number. However, if the data are sampled over a fairly evenly and finely spaced grid, then we might expect the observed estimate \hat{m}_1 to be fairly close to the true eDOT value δ_1 . We will see that the LMM slope estimate \hat{m}_1 does match $\hat{\delta}_1$ well in summarizing nonlinear growth when observations are evenly spaced over the study and there are no dropouts. However, the two estimators can diverge sharply either when scheduled observation times span the study but are unevenly spaced, or when there are dropouts.

4.2.2 eGaIT

The true eGaIT values are the definite integrals over $[0, 20]$:

$$\gamma_0 = \left(\frac{1}{2}(20 - 0)^2 \right)^{-1} \int_{t=0}^{20} \frac{1}{10} t dt = \frac{1}{10} = .10$$

and

$$\gamma_1 = \left(\frac{1}{2}(20-0)^2 \right)^{-1} \int_{t=0}^{20} \frac{1}{200} t^2 + \frac{-1}{5}t + \frac{5}{4} \left[1 - \cos \left(2\pi \frac{t}{20} \right) \right] dt = \frac{-1}{120} \approx -.0083$$

4.2.3 eDOT

It is straightforward to compute the true eDOT value over the interval $[0, 20]$ for each group by direct evaluation:

$$\delta_0 = \frac{g_0(20) - g_0(0)}{20 - 0} = \frac{7 - 5}{20} = .10$$

and

$$\delta_1 = \frac{g_1(20) - g_1(0)}{20 - 0} = \frac{3 - 5}{20} = -.10$$

so that in the linear case the control regimen has identical true values for eDOT and eGait ($\gamma_0 = .10$, $\delta_0 = .10$), while the treatment regimen has more discordant scores ($\gamma_1 = -.0083$ vs. $\delta_1 = -.10$).

4.3 Effects on the estimators of non-uniform measurement times.

In our simulation, animals were observed daily at time points $\{0, 1, \dots, 20\}$. We examine the performance of each of the LMM slope, eGait, and eDOT under four scenarios:

1. Models are fit with the full data set.
2. Data are artificially “thinned” before model fitting so that the number of observations early in the study is reduced. Specifically, remove data observed at times $\{1, 2, 4, 5, 6, 8, 9\}$ from consideration, so that the data set has times $\{0, 3, 7, 10, 11, 12, \dots, 20\}$. This is the ‘thin early’ scheme. The reason to include this scheme is to learn how much difference the scheduling of observation times within a study has on the different estimators, even when there are no dropouts so that each mouse’s tumor is measured at each time point.
3. Data are again thinned before model fitting, but now so that the number of observations *later* in the study is reduced. Specifically, remove data observed at times $\{11, 12, 14, 15, 16, 18, 19\}$, so that the data set has times $\{0, 1, \dots, 10, 13, 17, 20\}$. This is the ‘thin late’ scheme. As with the ‘thin early’ scheme, this will provide some insight into the effect on the estimators of unevenly scheduled observation times.
4. Finally, simulate a dropout mechanism dependent on tumor volume so that an animal is more likely to drop out when its tumor is large than when it is small. For each observation Y_{ijl} taken before the end of the study, let $\Delta x_{ijl} \equiv x_{ij,(l+1)} - x_{ijl}$ be the gap in time before the next scheduled observation. An animal drops out of the study before its next observation with probability $1 - e^{-\lambda(y_{ijl})\Delta x_{ijl}}$, i.e., the Poisson probability of having an event in the time interval following the current observation when the Poisson process rate is proportional to the hazard. Separately, an animal is automatically removed from the study if its observation y_{ijl} exceeds a protocol-defined threshold. The threshold was set at $\log(2001)$ in this simulation to mimic a hard threshold of 2000 mm^3 .

The tumor-volume-determined hazard function $\lambda(y_{ijl})$ was defined as $\lambda(y_{ijl}) = .01$ for tumors of size 500 mm^3 or less, then increasing linearly in log tumor volume up to $\lambda(y_{ijl}) = \log(2) \approx .693$ for tumors just below the limit of 2000 mm^3 . This implies that animals with tumors near 2000 mm^3 have a probability of nearly $1 - e^{-0.693} \approx .50$ of dropping out in each following day.

Each simulated study was retained conditional on both the control and treatment groups each having at least one mouse that did not drop out before the end of the study. This condition mimics the preclinical

reality that comparisons are made over the common intersection of ranges of recorded measurements. It was also a practical necessity to preserve a common basis of comparison across simulations, since the population parameter values for eGait and eDOT described above are computed for the interval [0, 20]. This requirement to have at least one animal that completed the interval resulted in re-simulating studies between 1% (for $n = 10$ per group) and 8% (for $n = 5$ per group) of the time.

In practice, of course, dropout probabilities likely will vary based not only on endpoint value, but also by treatment dose and schedule, mouse strain, tumor model, and likely other factors, so that inferring the dropout mechanism is beyond our current scope. Rather, this dropout model is intended merely to introduce into our simulation a mechanism by which animals with larger tumor volumes are more likely to be dropped mid-study, in order to investigate such a process' impacts on the different estimators.

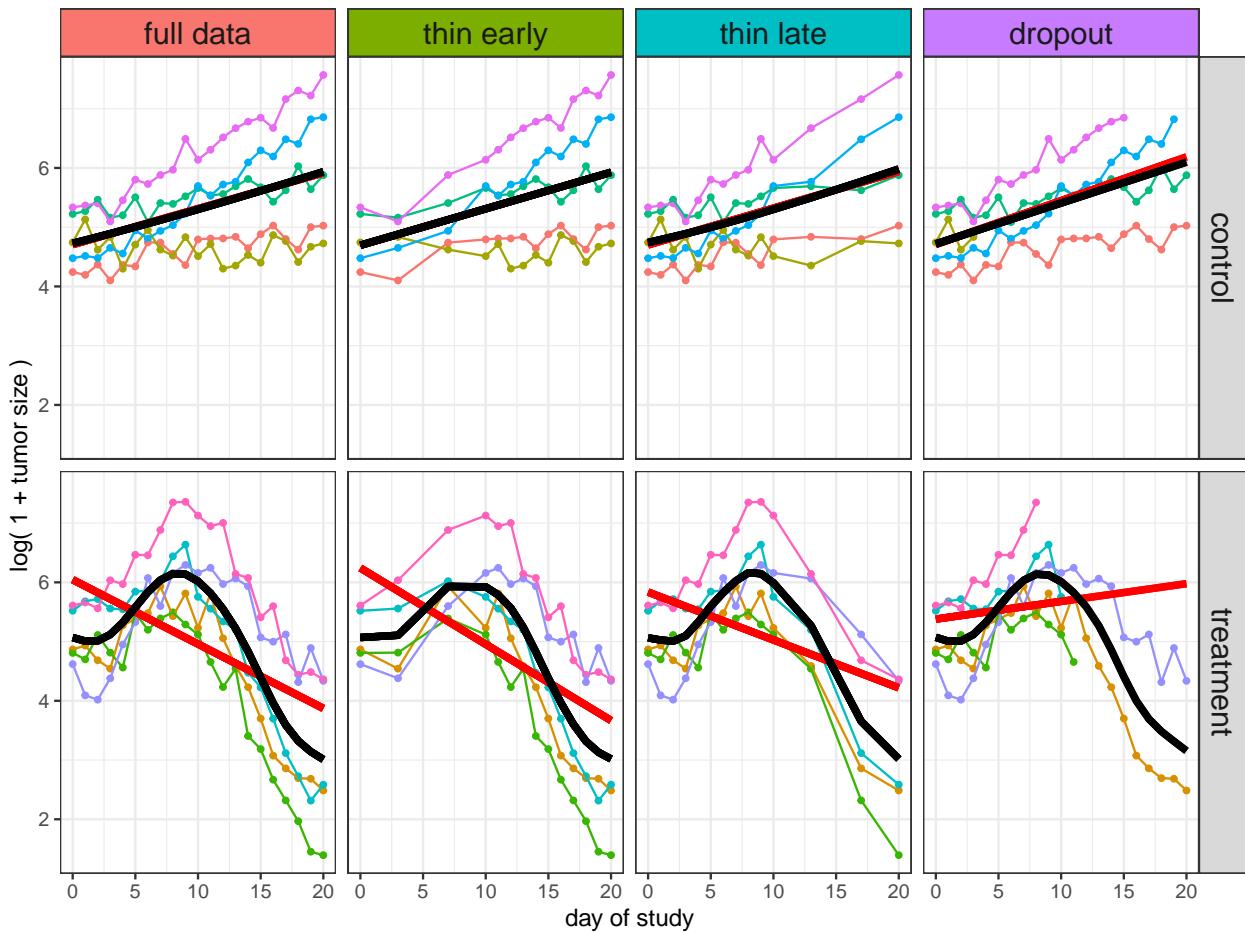
5 Results

For each simulation, LMM and GAMM models are fit, the LMM slope, eGait, and eDOT are computed, and model-based standard errors and 95% confidence intervals returned. We first present results for the point estimates under data thinning schemes and assess normality, then coverage for model-based confidence intervals.

5.1 Point estimates of LMM slope, eGait, and eDOT

Figure S.2 shows two-group data from one round of the simulation. Each column shows control and treatment in the upper and lower rows, respectively. From left to right, the columns show unredacted full data (first column), the same data with early time values thinned out (second), with later time values thinned out (third), and finally with endpoint-driven dropouts (fourth column). In each case, the LMM and GAMM models are fit and the resulting fixed effects curves overlaid on the data, with the LMM fit in red and the GAMM fit in black.

Figure S.2



Data from this two-group model were simulated 1000 times. In each round of the simulation, the four data filtering / dropout scenarios were applied and both a LMM and GAMM were fit to the data under each filtering scheme in turn (so that four LMMs and four GAMMs were fit in each round of the simulation). For each LMM, the LMM slope estimate and model-based 95% confidence interval were recorded for each group. Similarly, for each GAMM the statistics eGaIT and eDOT were computed along with their respective model-based 95% confidence intervals for each group.

Figure S.3 shows the 1000 point estimates for each of the LMM slope, eGaIT, and eDOT, for the control and treatment groups with 5 animals per group under each filtering scheme. The true (i.e., analytically calculated) parameter value in each case is denoted by a horizontal brown bar. The LMM estimates in the treatment group do not have such a bar included because the slope is undefined for a nonlinear model.

Setting aside LMM slopes for nonlinear treatment, the scatterplots in S.3 indicate that sampling distributions for the estimators are well-centered on their true values and the violin plot estimates show the distributions to be at least close to Gaussian (more on this below). The eGaIT and eDOT statistics are slightly biased in most scenarios, but the bias in all cases we evaluated is less than 10% of the standard deviation of the parameter estimates (often much less), so that the estimator variance is the dominant factor limiting estimator inference.

The LMM slope estimates for the nonlinear treatment data (the red points across the four lower panels of S.3) match closely with the eDOT statistics (i.e., the average first derivatives over the study interval) when the observation times are evenly spaced as in the `full data` scenario, but the sampling distribution of LMM slopes shift visibly downwards or upwards when the observation times are thinned in the early or late half of the study, respectively. When dropouts are present, the LMM slope is both higher and more variable. The intuition underlying this is that with dropouts, data from the descending second half of the growth curve are

underrepresented, while the (on-average) narrower span of observation times increases the slope estimate variability. In contrast, the GAMM spline summary measures are relatively unaffected by unevenly scheduled observation times or dropouts, so long as the data are sufficient to discern the shape of the underlying curve reasonably well, which they are here.

Figure S.3

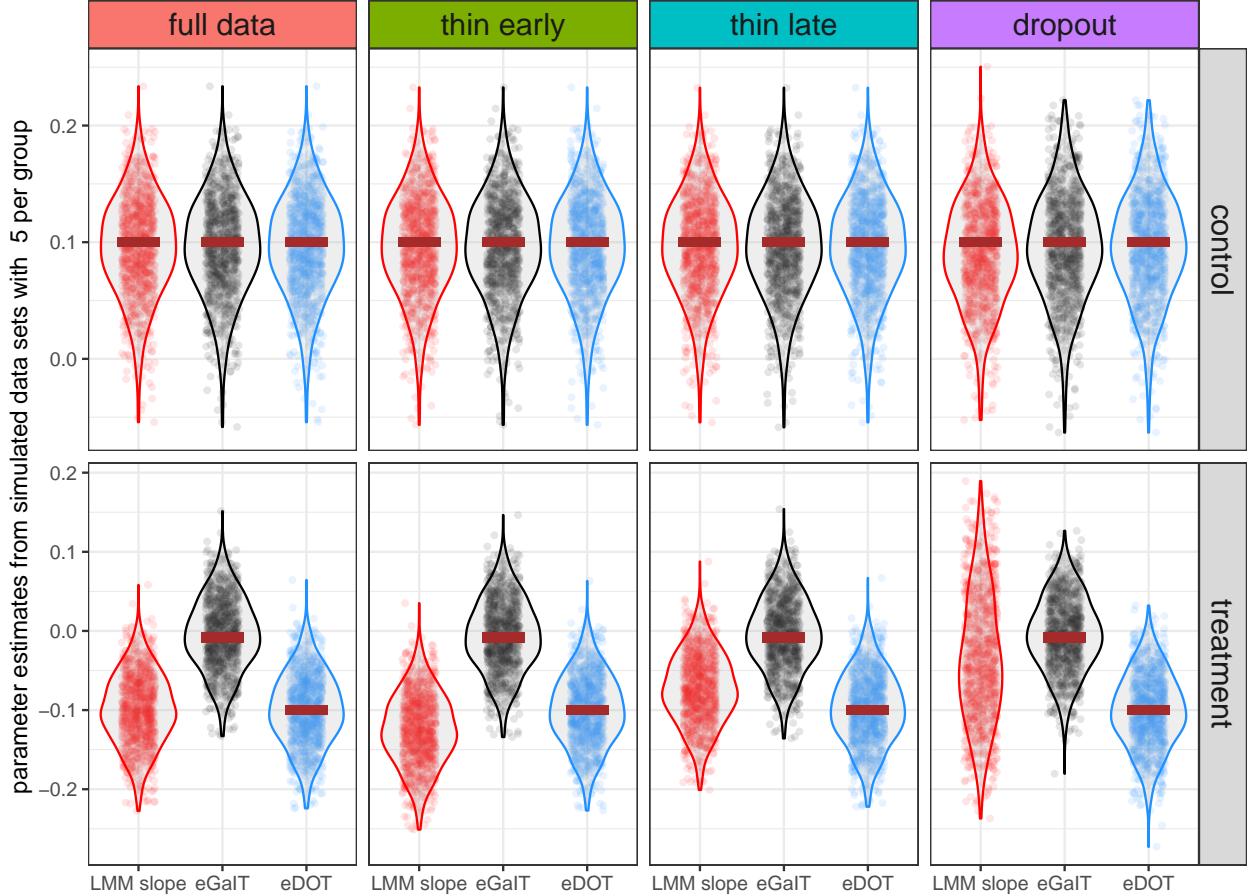


Figure S.4 shows the same parameter estimate points as Figure S.3, but now plotted against Gaussian quantiles so that points drawn from a normal distribution are expected to fall along a straight line. Perhaps surprisingly, deviation from linearity for eGaIT and eDOT in these quantile plots is greatest for estimates fit to the control data with a linear growth trend, while there is no evidence of non-normality (either visual or as assessed by a Shapiro-Wilk test) for eGaIT or eDOT estimates summarizing the nonlinear trend of the treatment regimen.

Figure S.5 shows the analogous quantile plots for simulations with 10 animals per group instead of 5 per group. The parameter estimates based on 10 animals per group do not show evidence of non-normality except in the case of the LMM slope estimates from the nonlinear treatment group (the points in red in the lower right panel). Note that since the LMM and GAMM both estimate common variance components across groups, the total number of animals across the full study (i.e., 10 or 20 per two-group study rather than 5 or 10 per group, respectively) will likely be more relevant for large-sample normality.

We hypothesize that the eGaIT and eDOT sampling distributions' heightened fidelity to the normal distribution in the nonlinear compared to the linear case may owe to the fact (discussed earlier) that both are linear combinations $c^T \hat{\beta}$ of the estimated spline basis coefficients. In the nonlinear case, there are many non-zero coefficients estimated for their respective spline basis functions in order to approximate the growth curve, with eGaIT or eDOT then being a weighted combination of those coefficients. While the same is technically true in the linear case, a linear term is always one of the basis vectors. Hence, when the spline looks linear, there

is effectively only one non-zero coefficient, so that the weighted averaging inherent in a linear combination no longer nudges the sampling distributions of eGaIT or eDOT toward normality. We extracted and examined the spline coefficient estimates from both cases and found that the sampling distributions of the coefficients across 1000 simulations were modestly non-normal, even though their eGaIT and eDOT summaries are normal in the nonlinear case.

Figure S.4

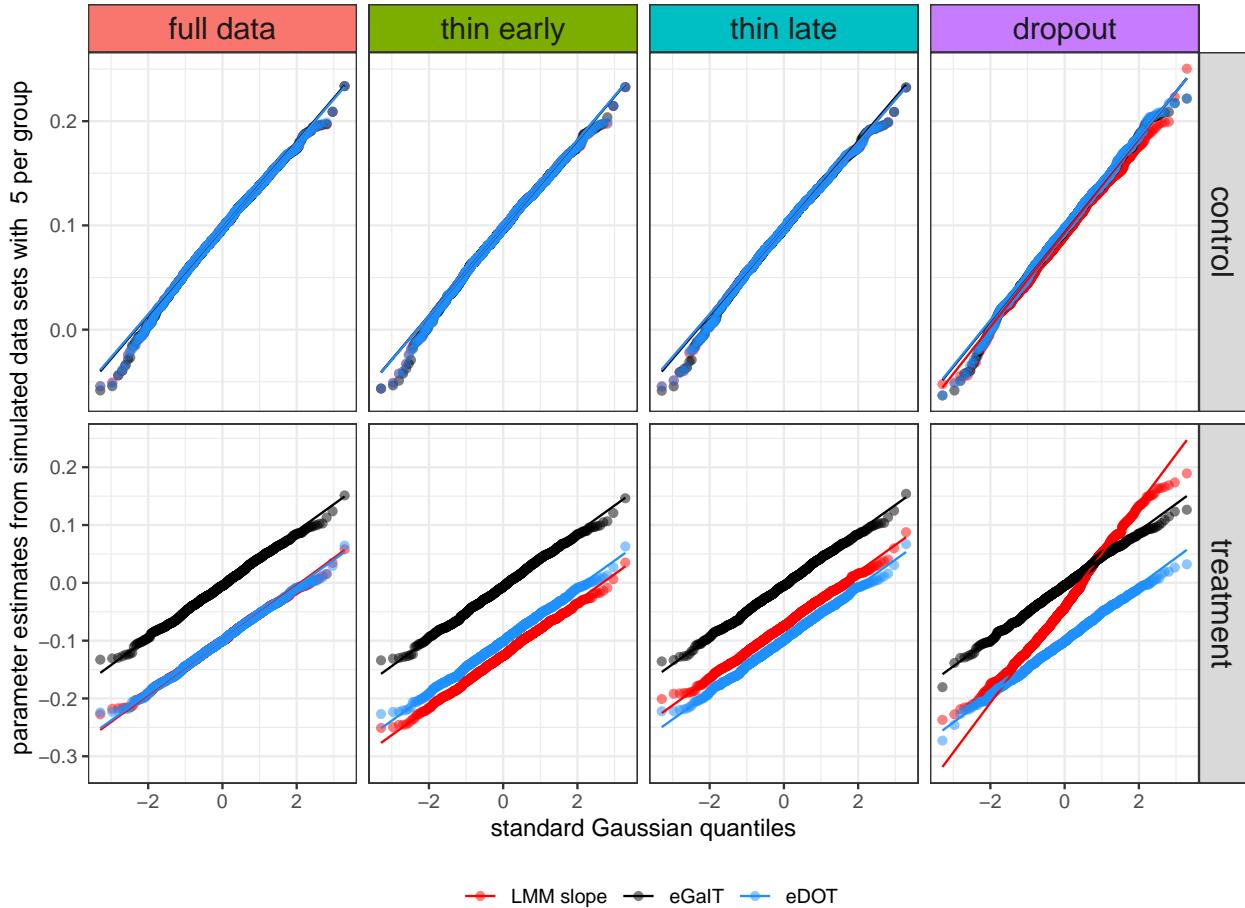
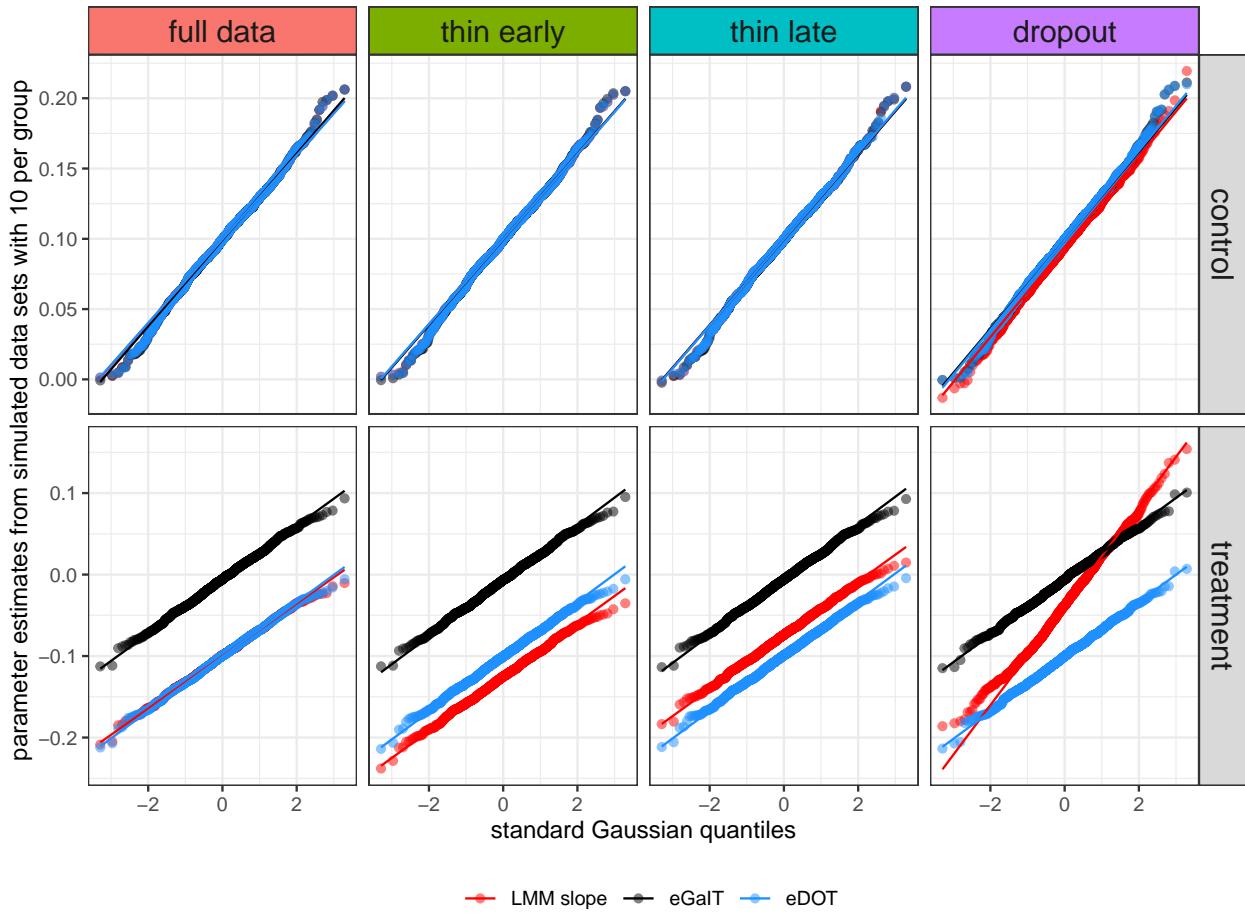


Figure S.5



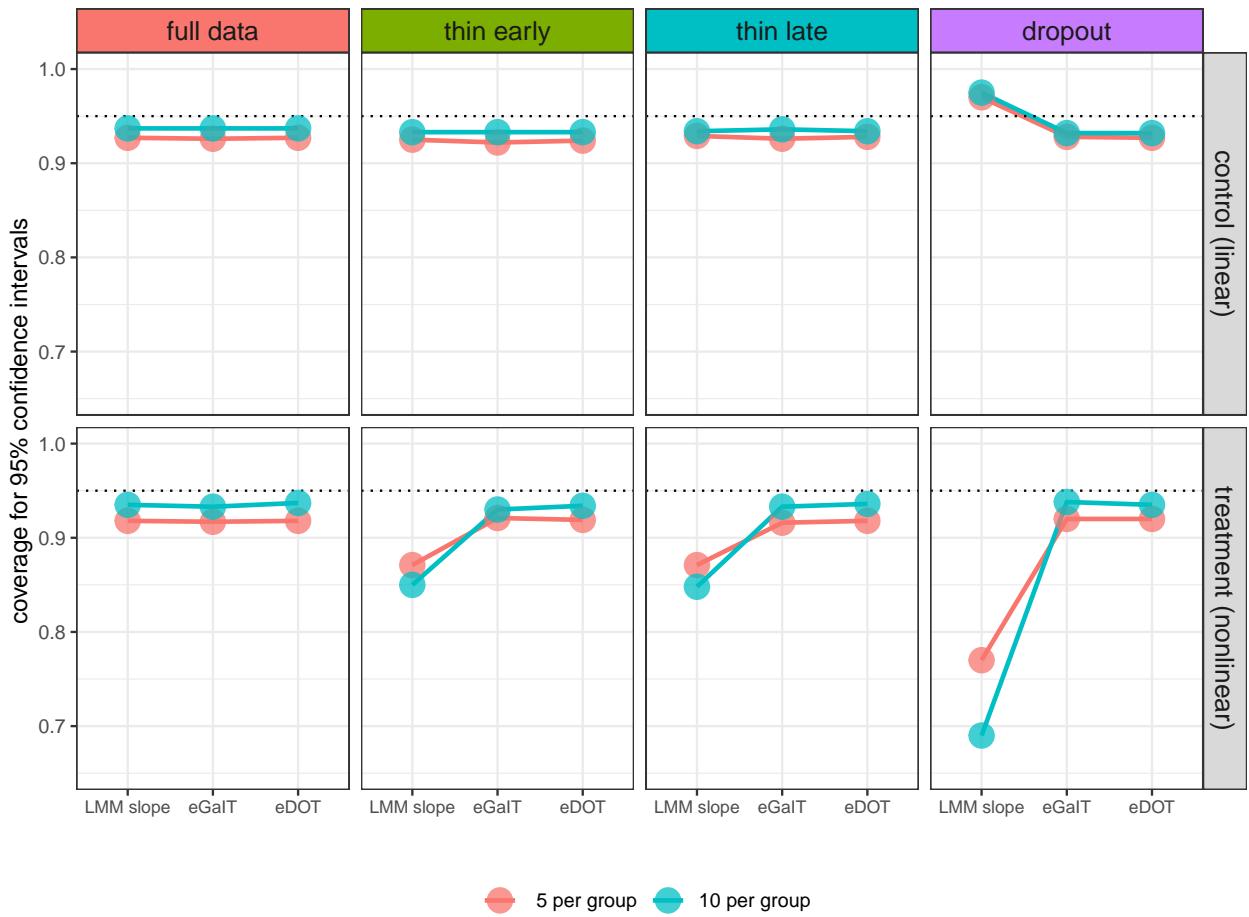
5.2 Confidence interval coverage

For each of the LMM slope, eGaIT, and eDOT estimates fit from simulated data, model-based standard errors (SEs) and 95% confidence intervals were returned, and interval coverage of the true parameter value assessed.

In the case of the LMM slope for the nonlinear treatment regimen (which, by definition, has no constant slope), we check the coverage of the true eDOT parameter values of $\delta_1 = -0.10$ (which is, recall, the time-averaged slope over the interval) with the rationale that this is what many users would regard as a reasonable target for a linear summary of a nonlinear curve.

Figure S.6 shows the fraction of times that the 95% confidence intervals (CIs) covered their nominal values. The CIs for eGaIT and eDOT are consistently anti-conservative but fairly close to the correct coverage, with roughly 92% coverage for a sample size of 5 animals per group to around 93% coverage with a sample size of 10 animals per group. We speculate that their coverages falling below 95% may owe both to the sample size and the small biases in the eGaIT and eDOT point estimates, mentioned earlier. Coverage for the LMM slope was similar to that of eGaIT and eDOT for the full data with equally spaced observations, but diverged markedly in scenarios for which the estimator is biased or especially noisy (i.e., when dropouts are present, or simply unevenly spaced observations in the nonlinear scenario).

Figure S.6



6 Discussion

In the log-linear tumor volume growth scenario represented by $g_0(t)$, all three estimators are usually quite similar to each other, which is expected since eDOT and eGaIT essentially reduce to the LMM slope in the linear case. The too-conservative coverage by the LMM slope CIs in the dropout scenario (upper right panel of S.6) stems from the fact that mixed models (both LMM and GAMM) estimate joint variance components across the control and treatment group. Data with both nonlinearity and dropouts tends to inflate the variance estimates fit from simple linear models as in the LMM, which in turn leads to over-coverage for the control group (i.e., for the control group the LMM slope point estimate is usually close to correct, but its CI is too wide because its CI is based in part on variance component estimates shared with the LMM slope fit to the *nonlinear* treatment group).

In the nonlinear tumor volume growth scenario represented by $g_1(t)$, the sampling distributions of estimates based on data sets with and without dropouts (shown by the violin plots in Figure S.3) indicate that the eDOT and eGaIT estimates are stable across scenarios, including the joint presence of nonlinearity and dropouts. In contrast, the sampling distribution of LMM slopes fluctuates depending on whether observations' times are unevenly scheduled or dropouts are introduced.

By examining Gaussian quantile plots of these estimates in Figure S.4 and S.5, we confirm that the eGaIT and eDOT estimators' sampling distributions are well approximated by normal distributions for the case of ten animals per group. The distributions depart somewhat from normality in the case of linear growth with five animals per group (here they look essentially like the LMM slope estimates), though fare better with five animals per group in the case of nonlinear growth. In all cases, there is not a noticeable effect in Figure

S.6 on coverages of their model-based confidence intervals, which are slightly below the nominal level but consistent across scenarios.

Taken together, these results indicate that in this admittedly limited simulation, the LMM slope can be capricious in the presence of nonlinearity and dropouts while the proposed GAMM summary statistics eGaIT and eDOT are relatively dependable in recapitulating their true values (though with slightly less than nominal coverage) making them usefully robust generalizations of the slope to summarize efficacy in longitudinal experiments.

7 References

1. Wahba G. Bayesian "confidence intervals" for the cross-validated smoothing spline. *Journal of the Royal Statistical Society Series B (Methodological)*. JSTOR; 1983;133–50.
2. Wahba G. A comparison of gcv and gml for choosing the smoothing parameter in the generalized spline smoothing problem. *The Annals of Statistics*. JSTOR; 1985;1378–402.
3. Silverman BW. Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society Series B (Methodological)*. JSTOR; 1985;1–52.
4. Friedman JH, Silverman BW. Flexible parsimonious smoothing and additive modeling. *Technometrics*. Taylor & Francis; 1989;31:3–21.
5. Polansky L, Robbins MM. Generalized additive mixed models for disentangling long-term trends, local anomalies, and seasonality in fruit tree phenology. *Ecology and evolution*. Wiley Online Library; 2013;3:3141–51.
6. Wood SN, Li Z, Shaddick G, Augustin NH. Generalized additive models for gigadata: Modeling the uk black smoke network daily data. *Journal of the American Statistical Association*. Taylor & Francis; 2017;112:1199–210.
7. Baayen RH, Rij J van, Cat C de, Wood S. Autocorrelated errors in experimental data in the language sciences: Some solutions offered by generalized additive mixed models. *Mixed-effects regression models in linguistics*. Springer; 2018. pages 49–69.
8. Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. Wiley Online Library; 2011;73:3–36.
9. Perperoglou A, Sauerbrei W, Abrahamowicz M, Schmid M. A review of spline function procedures in r. *BMC medical research methodology*. BioMed Central; 2019;19:46.
10. Wood SN. Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. Wiley Online Library; 2003;65:95–114.
11. Wood S. *Generalized additive models: An introduction with r*. 2nd ed. Chapman; Hall/CRC; 2017.
12. Simpson GL. Modelling palaeoecological time series using generalised additive models. *Frontiers in Ecology and Evolution*. Frontiers; 2018;6:149.
13. Wood S, Scheipl F. Gamm4: Generalized additive mixed models using 'mgcv' and 'lme4' [Internet]. 2016. Available from: <https://CRAN.R-project.org/package=gamm4>
14. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. 2015;67:1–48.
15. Forrest WF, Alicke B, Mayba O, Osinska M, Jakubczak M, Piatkowski P, et al. Generalized additive mixed modeling of longitudinal tumor growth reduces bias and improves decision making in translational oncology. (in revisions for resubmission). 2020.

8 Appendix

8.1 Appendix: Normalizing constants for spline summary statistics

Over an interval $[a, b]$, the spline summary statistics eDOT and eGaIT (i.e., δ and γ , respectively) are normalized by $b - a$ and by $\frac{1}{2}(b - a)^2$, respectively, so that their values will match closely with the slope in the linear case. Rationales for these constants are explained here.

8.1.1 eDOT

The choice of $(b - a)$ as a constant of normalization 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 taking a first derivative with respect to t of each side yields

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

Integrating both sides over $[a, b]$ then results in

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

so that normalizing to study length will recapture the slope $g^{(1)}(a)$ as the spline approaches a straight line, i.e., as $g^{(k)}(a) \downarrow 0, k \geq 2$:

$$\frac{1}{b - a} \int_a^b g^{(1)}(t) dt = \frac{g(b) - g(a)}{b - a} \equiv \delta_{[a,b]} \approx g^{(1)}(a)$$

8.1.2 eGaIT

The choice of $\frac{1}{2}(b - a)^2$ as a constant of normalization for γ is similarly motivated by again 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 area under the curve 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 γ converges smoothly to the slope as the spline straightens sufficiently towards a line, i.e., as $g^{(k)}(a) \downarrow 0, k \geq 2$:

$$\gamma \equiv \frac{1}{\frac{1}{2}(b-a)^2} \int_a^b g(t) - g(a) dt \approx g^{(1)}(a)$$

A more geometric intuition for $\frac{1}{2}(b-a)^2$ follows from the observation that insofar as a group's growth function $g(t)$ is roughly log-linear with growth rate m , then its estimated trajectory from study time $t = a$ through $t = b$ traces out a right triangle with base of length $b-a$ and height equal to $m(b-a)$. Since the area of this triangle (i.e., the definite integral) is $\frac{1}{2}m(b-a)^2$ our choice of normalization ensures that γ mimics the slope when growth is fairly log-linear, in which case either estimate can be interpreted as an average log fold change per unit of time.

9 Figure legends

Figure S.1: Underlying log tumor volume curves from the two growth models $g_0(t)$ and $g_1(t)$. The control curve $g_0(t)$ is linear when tumor volume is presented on the log scale, while the treatment curve $g_1(t)$ is nonlinear.

Figure S.2: One of 1000 simulated data sets each with 5 animals per group showing the $g_0(t)$ control group data along the upper row and $g_1(t)$ treatment group data along the lower row. The first column shows all of the simulated data. The second column shows the same data but with observations at times $\{1, 2, 4, 5, 6, 8, 9\}$ removed (this is the ‘thin early’ scenario). The third column shows the data with observations at times $\{11, 12, 14, 15, 16, 18, 19\}$ removed (this is the ‘thin late’ scenario). Finally, the fourth column shows the data with dropouts determined by the endpoint-size dependent dropout scheme, in which the mice with the largest tumors are disproportionately more likely to drop out.

Figure S.3: Parameter estimates over 1000 simulations with 5 animals per group for LMM slope, eGait, and eDOT under each of four data thinning schemes. The LMM slope fit to nonlinear data (lower row of panels) has a distribution that shifts noticeably based on when data are observed.

Figure S.4: Gaussian quantile plots for Parameter estimates over 1000 simulations with 5 animals per group for LMM slope, eGait, and eDOT under each of four data thinning schemes. Point estimators are fairly well-described by the normal distribution, but with some departures visible in the fits to the control data, and for the LMM slope fit to the nonlinear treatment data.

Figure S.5: Gaussian quantile plots for Parameter estimates over 1000 simulations with 10 animals per group for LMM slope, eGait, and eDOT under each of four data thinning schemes. Point estimators are well-described by the normal distribution, but with some departures visible for the LMM slope fit to the nonlinear treatment data.

Figure S.6: Coverage of 1000 model-based 95% confidence intervals (CI) based on $n = 5$ and $n = 10$ animals per group for LMM slope, eGait, and eDOT under each of four data thinning schemes. Coverage was slightly below the nominal value expected for the eGait and eDOT estimators across scenarios. LMM slope estimate coverage in the nonlinear treatment scenario tests whether the CI for the slope covers the true eDOT value of -.10.