Published OnlineFirst June 27, 2012; DOI: 10.1158/1078-0432.CCR-11-3215

Cancer Therapy: Preclinical

Clinical Cancer Research

Improved Statistical Modeling of Tumor Growth and Treatment Effect in Preclinical Animal Studies with Highly Heterogeneous Responses *In Vivo*

Teemu D. Laajala[1,6], Jukka Corander[8], Niina M. Saarinen[4,5], Katja Mäkelä[2,4,5], Saija Savolainen[2,5], Mari I. Suominen[7], Esa Alhoniemi[7], Sari Mäkelä[4,5,3], Matti Poutanen[2,5,10], and Tero Aittokallio[1,6,9]

Abstract

Purpose: Preclinical tumor growth experiments often result in heterogeneous datasets that include growing, regressing, or stable growth profiles in the treatment and control groups. Such confounding intertumor variability may mask the true treatment effects especially when less aggressive treatment alternatives are being evaluated.

Experimental design: We developed a statistical modeling approach in which the growing and poorly growing tumor categories were automatically detected by means of an expectation-maximization algorithm coupled within a mixed-effects modeling framework. The framework is implemented and distributed as an R package, which enables model estimation and statistical inference, as well as statistical power and precision analyses.

Results: When applied to four tumor growth experiments, the modeling framework was shown to (i) improve the detection of subtle treatment effects in the presence of high within-group tumor variability; (ii) reveal hidden tumor subgroups associated with established or novel biomarkers, such as ERβ expression in a MCF-7 breast cancer model, which remained undetected with standard statistical analysis; (iii) provide guidance on the selection of sufficient sample sizes and most informative treatment periods; and (iv) offer flexibility to various cancer models, experimental designs, and treatment options. Model-based testing of treatment effect on the tumor growth rate (or slope) was shown as particularly informative in the preclinical assessment of treatment alternatives based on dietary interventions.

Conclusions: In general, the modeling framework enables identification of such biologically significant differences in tumor growth profiles that would have gone undetected or had required considerably higher number of animals when using traditional statistical methods. *Clin Cancer Res; 18(16); 4385–96. ©2012 AACR.*

Introduction

Preclinical tumor growth studies using animal models have a fundamental role in anticancer drug development. Experimental cancer models in mice and rats include, among others, implanting human tumor cells into immunocompromised animals (xenograft models) or inducing tumor-promoting mutations in rodents using carcinogens such as 7,12-dimethylbenz(*a*)anthracene (DMBA). Regardless of the model type, the typical experimental design involves dividing the animals into the treatment groups (representing different doses or treatment combinations), and monitoring the relative effects of the treatments on tumor growth, in comparison with the control group (no treatment). The tumor growth is typically measured at a number of time intervals until the animals die, become moribund, or reach a planned time of sacrifice (1).

Despite careful control of the experiments, the longitudinal tumor growth measurements reflect multiple sources of both biologic and experimental variation that may severely confound the actual treatment responses. Along with measurement noise, additional experimental challenges include missing data points due to animal morbidity, mortality, and quantitation limits, as well as very aggressively growing outlying profiles. Such experimental variation can be compensated to some degree by increasing the number of animals and tumors analyzed. However, due to economical and practical reasons, most

Authors' Affiliations: [1]Department of Mathematics, [2]Departments of Physiology and [3]Cell Biology and Anatomy, Institute of Biomedicine, [4]Functional Foods Forum, [5]Turku Center for Disease Modeling, University of Turku; [6]Turku Centre for Biotechnology; [7]Pharmatest Services Ltd, Turku; [8]Department of Mathematics and Statistics, [9]Institute for Molecular Medicine (FIMM), University of Helsinki, Finland; and [10]Institute of Medicine, The Sahlgrenska Academy, Gothenburg University, Gothenburg, Sweden

Note: Supplementary data for this article are available at Clinical Cancer Research Online (clincancerres.aacrjournals.org).

Current address for S. Savolainen: Orion Pharma Ltd, Turku, Finland.

Corresponding Author: Tero Aittokallio, Institute for Molecular Medicine Finland, University of Helsinki, FI-00014, Finland. Phone: 358-50-318-2426; Fax: 358-9-191-25737; E-mail: tero.aittokallio@fimm.fi

doi: 10.1158/1078-0432.CCR-11-3215

Published OnlineFirst June 27, 2012; DOI: 10.1158/1078-0432.CCR-11-3215

Translational Relevance

Heterogeneous responses observed in many preclinical models of cancer treatment may lead to frequent false-negative results and therefore to ineffective translation of *in vivo* results to clinical trial designs. Using various preclinical animal models, cancer cell lines, and *in silico* simulations, we show here how modeling and exploring of different categories of tumor growth profiles can improve statistical testing and biologic understanding of treatment effects, especially when less aggressive treatment alternatives are being evaluated. Statistical power and precision analyses offer possibilities for further improving the design of the experimental protocols for preclinical assessment of cancer treatments. Taking into account the individual characteristics already in the preclinical stages should also help to propagate information on the intertumor variability to the subsequent clinical studies.

experiments are still being carried out on relatively small sample sizes including less than 10 tumors per group (1). Moreover, even when using genetically standardized and well-characterized animal strains, the experiments often represent substantial between-animal variability, which cannot be controlled simply by increasing the number of animals. Such confounding factors often result in hidden subgroups, which are not predefined but may associate with divergent treatment outcomes in terms of the growth profiles observed over the treatment period. Some tumors may grow aggressively in a treatment group, even if the same treatment inhibits the growth of other tumors, or some untreated tumors do not grow well or even completely regress in the control group (2–8).

The heterogeneous nature of the tumor growth profiles pose severe challenges to the statistical models that typically rely on the assumption that the groups being compared are relatively homogeneous. Many studies have used single end points, such as tumor volume at a prespecified time point or tumor doubling time, together with traditional statistical tests, such as *t* test and ANOVA, or their nonparametric counterparts (5–11). However, such univariate approaches often lead to suboptimal statistical power because of their ineffective use of the longitudinal growth patterns (1, 12). In contrast, repeated measures and regression models use the entire growth profiles and enable more systematic between-group comparisons through model parameters (1). In particular, mixed-effects models have become a convenient approach to model various experimental factors, such as treatment effects or base levels (fixed effects) while accounting for variation expressed by individual animals or tumors (random effects). This model family has successfully been used to analyze specific types of xenograft experiments or study questions (12–17). However, further challenges remain. In particular, the conventional model cannot detect subtle treatment effects in the presence of heterogeneous responses, due to unfeasible model estimation, resulting in skewness or multimodality in the random effects (18).

The present work introduces a novel modeling framework for in-depth statistical analysis of tumor growth experiments in which the underlying tumor heterogeneity is modeled by dividing the longitudinal growth profiles into growing and poorly growing categories within the treatment and control groups. The framework is based on well-established linear mixed-effects models enabling robust estimation and statistical inference of treatment effects through parameters such as tumor growth rates (slopes) or average tumor levels (offset). By means of such elemental parameters that are descriptive of both strong and more subtle modes of tumor growth inhibition, the modeling framework enables the investigator to address a range of questions relevant in many practical settings, such as the degree of dynamic treatment effect on the growth rates, the amount of tumor heterogeneity present in the given data, and how the experimental design should be modified to find significant treatment effects. To promote its widespread application in the future studies, we provide an easy-to-use R implementation with accompanying tools for model visualization and diagnostics. Using 4 tumor growth experiments as application use cases, we show here how the categorizing mixed-effects model enables the extraction of full information from these longitudinal profile datasets.

Materials and Methods

The model was applied to 4 tumor growth experiments, including prostate and breast cancer mouse xenograft models, a syngeneic mammary cancer model with 4T1 mouse mammary tumor cells, and a DMBA-induced mammary carcinoma in the rat. These experiments represent with a wide range of properties encountered in many treatment settings, including various treatment options and dosages (Table 1). Moreover, the experiments included designs with and without a designated target size that the tumors need to reach before treatment initiation. The designs differed also in the number of tumors per treatment group, a parameter, which is directly related to the power of detecting statistically significant treatment effects. Other experimental design parameters included diverse setups for treatment periods and sampling frequencies as well as different response readouts such as tumor volume or area. Importantly, 3 of the 4 experiments showed different degrees of intertumor heterogeneity in terms of evidence for within-group growing and poorly growing categories (Supplementary Fig. S1).

DMBA-induced mammary cancer model

Anticarcinogenic activity of the diet-derived lignan metaboline, enterolactone (ENL), was studied by applying a mammary cancer model in the rat (6) in which the mammary tumors were induced by the use of DMBA. The induction caused a varying number of tumors per animal (1–5 measurable tumors) and thus the total tumor

Published OnlineFirst June 27, 2012; DOI: 10.1158/1078-0432.CCR-11-3215

Table 1. Summary of the experimental datasets used in the present work

Experiment	DMBA case	MCF-7 case	LNCaP case	4T1 case
Strain	Female Sprague-Dawley rat	Female athymic nude mouse	Male athymic nude mouse	Female immunocompetent balb/c mouse
Cell line (source)		MCF-7 (human)	LNCaP (human)	4T1 (mouse)
Cancer model	Breast cancer, carcinogen	Breast cancer, xenograft	Prostate cancer, xenograft	Breast cancer, syngeneic
Treatment	ENL	LAR	DPN or ENL	Doxorubicin or cyclophosphamide
Dosage and route of administration	Daily 1 or 10 mg/kg *per os*	Daily 20 or 100 mg/kg *per os*	DPN 4.5 mg/60 days s.c. or ENL 100 mg/kg in feed	Doxorubicin: weekly 7.5 mg/kg; cyclophosphamide: 100 mg/kg at days 0, 2, and 4
Measurement frequency	Once a week	Once a week	Twice a week	Twice a week
Number of time points	9	6	11	6
Sample sizes	13 animals per group	Control (15), lower dose (20), higher dose (20) tumors	Control (12), DPN (10), ENL (8) tumors	8 tumors per group
Target size	No	20 $mm^{2,a}$	200 $mm^{3,b}$	No
Response readout	Total tumor volume per animal	Tumor area	Tumor volume	Tumor volume
Missing value proportions	Control 4% Low dose 4% High dose 0%	All groups 0%	Control 14% DPN 10% ENL 9%	Control 0% Doxorubicin 2% Cyclophosphamide 0%
Additional cell markers	Tumor histologic types	ERα, ERβ	PSA	Metastases in the lung and liver
Reference	6	19	Unpublished	21

[a]Treatment starting time defined by average tumor area
[b]Treatment starting time defined by individual tumor volume

volume per animal was used as the response readout (Table 1). Two different dosages of ENL (1 and 10 mg/kg *per os* by gavage) were introduced 9 weeks after the DMBA induction. Each of the treatment groups included both growing (growth profiles with positive slope) and poorly growing (horizontal profiles near zero volume) tumors; the lower dosage group also contained 2 outlier profiles (Supplementary Fig. S1A). All the profiles were used here in the statistical modeling.

Histologic classification of the tumors was carried out as described earlier (6). Briefly, the tumor contributing most to the total volume per animal was considered, as it was most often histologically analyzed and could be considered as most representative for the animal. Some of the tumors could not be analyzed due to issues related to tumor suppression, volume below detection accuracy, or quality of the sample. The histologic types of "poorly differentiated", "well differentiated", and "atrophic" included more than one tumor and these were used in the analyses.

MCF-7 breast cancer xenograft model

MCF-7 breast cancer xenografts were grown in ovariectomized athymic mice in the presence of estradiol (19). The antitumor activity of the dietary lignan, lariciresinol (LAR), was studied by applying 2 different dosages (20 or 100 mg/kg *per os* by gavage) of the compound, and the tumor growth was compared with mice treated with the vehicle only (Table 1). The tumor growth profiles were analyzed along with biomarkers, such as estrogen receptors α (ERα, ESR1) and β (ERβ, ESR2), to identify explanatory factors for the observed heterogeneous growth profiles (Supplementary Fig. S1B).

LNCaP prostate cancer xenograft model

This experiment studied the effects of a synthetic ERβ-selective agonist [DPN; 2,3-bis(4-hydroxyphenol)-propionitrile] and of a tissue-specific ER activator, diet-derived lignan metabolite (ENL) on the growth of the LNCaP prostate cancer xenografts in immunocompromised

Published OnlineFirst June 27, 2012; DOI: 10.1158/1078-0432.CCR-11-3215

mice (Athymic Nude-Fown 1 nu, Harlan). The cells (2×10^6 cells/200 μL medium/Matrigel) were subcutaneously inoculated into 5- to 6-week-old male mice. DPN was administered as pellets (4.5 mg for 60 days, Innovative Research of America). The mice in both control and treatment groups were fed purified control diet (AIN-93G; ref. 20). ENL was provided within a special diet including 100 ppm of the compound. The tumors were palpated twice a week, and the treatment was commenced once a tumor reached the target volume of 200 mm^3. To maximize the number of tumors, the growth period was allowed to reach the target volume level within 4 to 6 weeks. Because the number of tumors in the experiment remained relatively small, a maximal number of the short and outlier profiles, which often are filtered out in standard analyses, were included in the statistical analysis of the heterogeneous dataset (Supplementary Fig. S1C). In addition to the tumor size, serum prostate-specific antigen (PSA), a known prostate cancer biomarker, was measured at sacrifice (Table 1).

4T1 syngeneic mammary cancer model

Mouse mammary adenocarcinoma 4T1-cells (American Type Culture Collection) were inoculated into the thoracic mammary fat-pads of 6-week-old female immunocompetent Balb/c mice (Harlan Laboratories Inc.; ref. 21). Two established drugs were used for the treatments (Table 1): doxorubicin (22) and cyclophosphamide (23). The drug treatments were started 6 days after the inoculation of the cells. Doxorubicin (Doxorubicin Ebewe; Ebewe Pharma GmbH) was administered 7.5 mg/kg once a week and cyclophosphamide (Sendoxan, Baxter) 100 mg/kg was administered at days 0, 2, and 4 since the beginning of the treatment. The tumor growth profiles showed very homogeneous patterns within each of the treatment groups (Supplementary Fig. S1D), possibly due to the host environment being native to the 4T1 cancer cell line (24).

The categorizing mixed-effects model

The mixed-effects models have a number of advantages in the statistical analysis of tumor growth profiles. First, the whole longitudinal growth profile, with possible missing data points, can be used in the model estimation and parametric inference thereby avoiding the need for selecting predefined endpoints or *ad hoc* imputation of missing values. Second, the random effects give flexibility for the model to take into account individual tumor- and animal-specific variation that originates from the given experimental setup and data. We extended the standard model and developed a novel, hierarchical mixed-effects model, which learns the growing and poorly growing tumor categories in a given set of longitudinal tumor growth profiles. The categorizing mixed-effects model is conceptually formulated as:

$$\begin{aligned}\text{Tumor response} = {} & b_1 + b_2 \times \text{Treatment} + b_3 \\ & \times \text{Time point} \times \text{Growth} + b_4 \\ & \times \text{Treatment} \times \text{Time point} \times \text{Growth} \\ & + u_{1,T} + u_{2,T} \times \text{Time point} \quad \text{(Model 1)}\end{aligned}$$

Here, the binary treatment covariate indicates the control and treatment groups and time point indicates the discrete measurement time points (Supplementary Table S1). The binary growth covariate is used to distinguish between the growing and poorly growing tumor categories. The terms b_i represent the model's fixed effects accounting for factors such as the base level tumor size (b_1), treatment-induced shift in the average tumor levels over the timepoints (offset, b_2), overall growth rate of those tumors categorized as growing (b_3), and treatment-induced difference in the growth rate of the growing tumors (slope effect, b_4). The random effects $u_{1,T}$ and $u_{2,T}$ represent variation specific to an individual tumor T. The full mathematical model formulation and details of its estimation, inference, and validation are given in Supplementary Methods.

Testing for the treatment-effects is done through the parameter estimates from the fitted categorizing model (Fig. 1A). The slope effect term b_4 evaluates time-dependent changes in the relative tumor growth rate per time

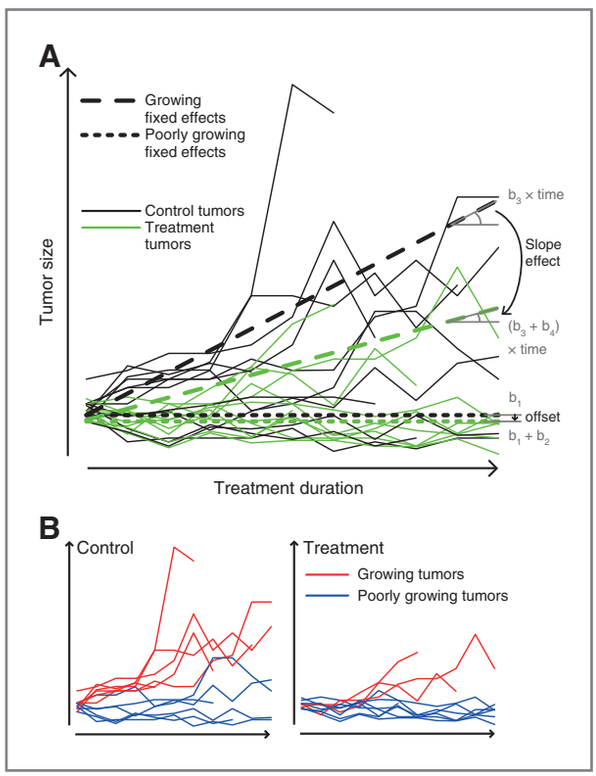

Figure 1. Schematic illustration of the treatment effect assessment in the LNCaP DPN experiment. A, fixed effects of the categorizing mixed-effects model are estimated from the data (b_1-b_4). The slope effect evaluates a treatment-induced and time point–dependent decrease in the growth rates of the growing tumors, whereas the offset term evaluates a treatment-induced shift in the horizontal tumor levels over all the time points and tumors. B, once the growing and poorly growing categories have been found by the model, the category labels are tested against the treatment labels, hence enabling evaluation of potentially more complex growth inhibiting treatment effects that may not be directly reflected in the offset or slope effects (here $P = 0.415$, Fisher exact test; Table 3).

Published OnlineFirst June 27, 2012; DOI: 10.1158/1078-0432.CCR-11-3215

unit in tumors categorized as growing. The slope effect therefore captures also a subtle suppressive treatment-effect relative to the overall growth rate b_3. An effective growth inhibition rate was defined as $|b_4/b_3|$. The offset term b_2 in turn evaluates more dramatic changes in the horizontal base level profiles of the tumors in those studies with a designated target size; otherwise, the terms b_1 and b_2 are set equal to zero. Because these terms do not account for the dynamic changes in the treated or control profiles, the offset term effectively captures the average treatment response in the poorly growing tumors over the entire treatment duration.

We implemented a novel clustering method based on the expectation-maximization (EM) algorithm for categorizing the tumor profiles into the growing and poorly growing subgroups (Supplementary Fig. S2). The model fitting was done using the restricted maximum likelihood (REML) estimation in the lme4 package (25) within the R statistical software (26). The statistical significance of the treatment-specific fixed effects was assessed through Markov-Chain Monte Carlo (MCMC) simulation (27). The full details of the implementation of the modeling framework are given in Supplementary Methods. The source code of the implemented R package, named XenoCat, is freely available (28).

Post hoc statistical analyses

After the growth categories were detected from the fitted model, 2-sided Fisher exact test was used to assess whether the found categorization into the growing and poorly growing subcategories can be explained by the proportion of tumors from the control and treatment groups (Fig. 1B). Significant overrepresentation of the treated tumors in the poorly growing category is indicative of such treatment effect that inhibits the tumor growth but may not be directly reflected in the fixed effect terms of the model. Hence, the offset and slope effect terms, together with the *post hoc* analysis of the detected growth categories using the Fisher exact test, can be used to draw conclusions on the treatment effects and underlying mechanisms of action.

In addition to the treatment labels, other external biologic and experimental explanatory factors for the growing and poorly growing categories were subsequently tested. For discrete explanatory factors, such as the histologic tumor classification, the Fisher exact test was used to assess the association between the tumor growth labels and the histologic classes. For normally distributed continuous factors, such as the ERβ positivity, the Welch 2-sample unpaired t test was used to evaluate the difference in the ERβ expression between the 2 growth categories. In case the Shapiro–Wilk normality test null hypothesis was rejected, the Wilcoxon rank-sum test was used instead as a nonparametric alternative.

Power, precision, and sample size estimation

Comparisons between different experimental designs and modeling setups were carried out to provide further model-guided information on their operation and suggestions for future improvements. The comparisons were based on parameters, such as the number of tumors and/or timepoints, which were investigated in relation to the calculated statistical study power, defined as the probability of detecting a statistically significant treatment effect, provided that the effect is truly present and that the model is correct. Estimation of the sample size N that is needed to achieve a given statistical power was based on simulated data generated according to the model fit (29). Furthermore, a precision analysis was implemented using the modeling framework to give guidance on the most informative time periods. Precision here means the reciprocal of the variance of the test statistic, given the estimated model and the experimental design (30). A general overview of the modeling workflow is available in Supplementary Fig. S3.

Results

An efficient implementation of the statistical modeling framework was developed and distributed as an open-source R package, named XenoCat, with accompanying user instructions (28). Here, the framework was applied to 4 case studies and the results from the categorizing mixed-effects model were compared with those obtained using the conventional mixed-effects model in terms of statistical inference, power, precision, and suggested sample size. The conventional noncategorizing mixed-effects model is a special case of the Model 1, in which the growth covariate is omitted (i.e., set to unity).

DMBA case

Estimation of the categorizing mixed-effects model in the DMBA experiment illustrates how the model can effectively describe the growing and poorly growing tumor subcategories within the treatment and control groups (Fig. 2). By taking into account such tumor growth heterogeneity, the categorizing model gave highly significant treatment effect on the slope effect term consistently both in the in low-dose and the high-dose groups ($P = 7.4 \times 10^{-5}$, $|b_4/b_3| = 40\%$ and $P = 3.8 \times 10^{-5}$, $|b_4/b_3| = 49\%$; Table 2). The subtle suppressive effect of the dietary intervention (ENL treatment) on the growth rate was missed by the conventional mixed-effects model even in the high dosage treatment group ($P > 0.05$). The increased sensitivity of the categorizing model is due to improved model fit, as indicated by the loss of skewness and multimodality in the distribution of the random slopes (Supplementary Fig. S4).

Because the identified growing and poorly growing categories could not be explained by the ENL treatment groups (Fig. 2D; Table 3), we searched for explanatory factors from the histologic analysis of the tumors. According to expectations, the tumors classified as "well differentiated" or "atrophic" were decreased in proportion in the growing tumor category consistently under both dosage levels, whereas the tumors classified as "poorly differentiated" were more abundant in the growing category (Table 3). Even if showing only a borderline statistical

Published OnlineFirst June 27, 2012; DOI: 10.1158/1078-0432.CCR-11-3215

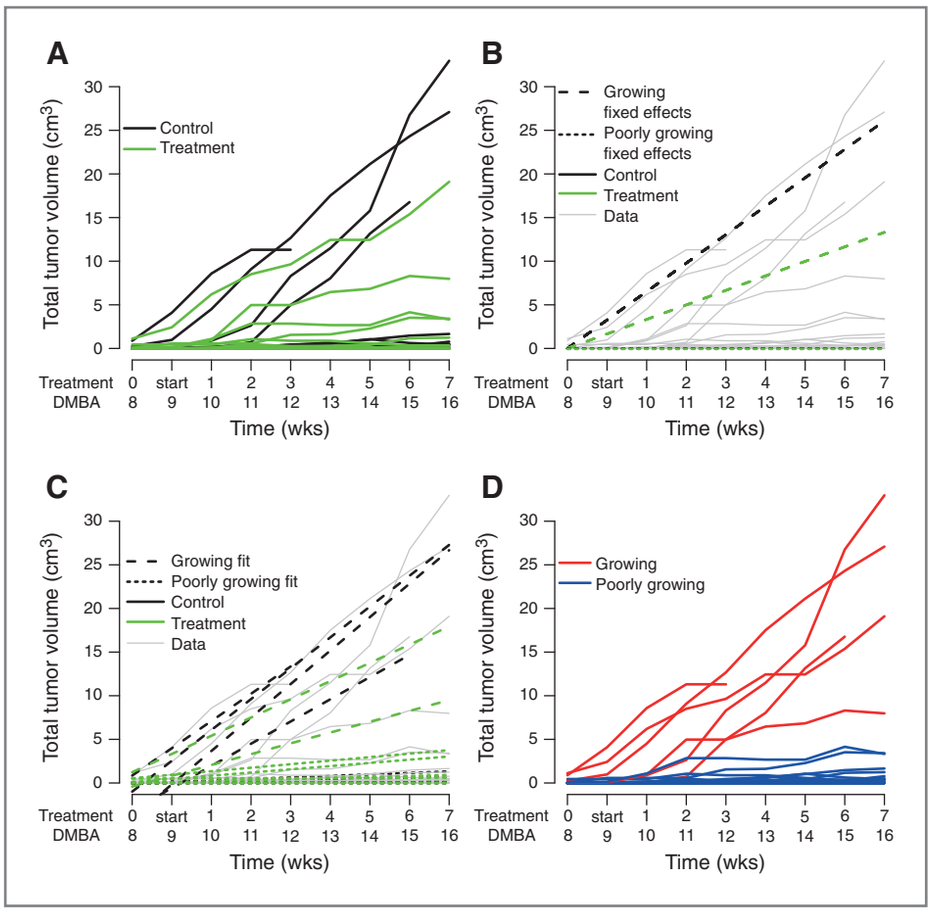

Figure 2. Operation of the categorizing modeling procedure in the DMBA ENL high-dose experiment. A, distinct growth patterns are evident in individual rats within both groups where some tumors grow aggressively, whereas others remain completely stabilized during the treatment period. B, the fixed effect fit for the population growth profiles show treatment-induced slope difference in the growing tumors ($P < 0.001$). C, the full model fit, where the individual variation is modeled using random-effects. D, The growing and poorly growing tumor categories found by the EM algorithm. In each panel, the 'DMBA' time scale depicts the time as tumor induction by the carcinogen DMBA, and the 'treatment' time scale depicts the time since the treatment initiation (start).

association ($P = 0.069$), the relative proportion of tumors in the histologic classes supported the existence and relevance of the 2 growth categories. In contrast, the association between the treatment groups and the histologic types was highly insignificant ($P = 0.955$) indicating that the treatment *per se* did not influence the differentiation process.

MCF-7 case

In the MCF-7 xenograft experiment, the effects of the dietary lignan LAR treatment were found insignificant both on the offset and slope terms (Table 2). However, even in the absence of statistically significant treatment effects, the growing and poorly growing tumor categories could be explained by the treatment groups under the high dosage LAR treatment (Fisher exact test, $P = 0.022$; Table 3), suggesting that the dietary lignan treatment successfully blocks a significant portion of tumors into the poorly growing category. Interestingly, the tumors in the growing and poorly growing categories were also different in terms of their measured ERβ levels in the high dosage group ($P = 0.008$; Table 3) indicating that ERβ inhibits tumor growth, as has been previously suggested on the basis of results obtained from other experimental breast cancer models (31).

LNCaP case

A xenograft study with LNCaP cells was analyzed in terms of possible treatment effects, and to provide guidance for a sufficient sample size and the most informative time periods to be used in further studies. The categorizing model showed, already in the present data, a statistically highly significant slope effect in response to the ENL treatment ($P = 0.001$, $|b_4/b_3| = 80\%$), and a slightly significant slope effect in response to the DPN treatment ($P = 0.037$, $|b_4/b_3| = 48\%$). Both of these effects were undetected by the conventional mixed-effects model ($P > 0.05$; Table 2). However, both model types captured well the target tumor volume of 200 mm^3 in their base level terms under both treatments ($P < 10^{-5}$), whereas the categorization emphasized the overall growth terms ($P < 10^{-5}$).

The measured PSA concentrations at sacrifice were significantly different between the tumors classified into the growing or poorly growing categories. According to expectations, the PSA levels were consistently higher in the growing category than in the poorly growing category both in response to the DPN ($P = 0.005$) and ENL ($P = 0.001$) treatments (Table 3). Interestingly, the PSA levels at sacrifice were similar in the control and treatment groups both on DPN and ENL ($P > 0.05$) indicating that factors other than

Published OnlineFirst June 27, 2012; DOI: 10.1158/1078-0432.CCR-11-3215

Table 2. Fixed effect estimates, confidence intervals and statistical significance

DMBA case	**Categorizing model**			**Noncategorizing model**		
ENL low dose	Estimate	HPD interval	*P*	Estimate	HPD interval	*P*
Overall growth b_3	3.12	[2.56 to 3.39]	<0.001	0.989	[0.34 to 1.54]	<0.01
Slope effect b_4	−1.25	[−1.74 to −0.703]	<0.001	−0.174	[−1.04 to 0.66]	0.659
ENL high dose						
Overall growth b_3	3.26	[2.78 to 3.47]	<0.001	1.02	[0.50 to 1.44]	<0.001
Slope effect b_4	−1.60	[−2.01 to −0.871]	<0.001	−0.681	[−1.28 to 0.04]	0.066
MCF-7 case						
LAR low dose						
Base level b_1	21.4	[16.9 to 25.8]	<0.001	21.0	[16.5 to 25.8]	<0.001
Offset b_2	−2.66	[−8.68 to 3.01]	0.359	−4.18	[−10.5 to 1.97]	0.183
Overall growth b_3	8.71	[6.92 to 10.6]	<0.001	8.37	[6.55 to 10.2]	<0.001
Slope effect b_4	−0.599	[−3.10 to 1.94]	0.619	−1.04	[−3.44 to 1.44]	0.398
LAR high dose						
Base level b_1	21.4	[17.0 to 25.7]	<0.001	21.0	[16.5 to 25.7]	<0.001
Offset b_2	1.05	[−5.00 to 6.28]	0.802	−0.945	[−7.12 to 5.08]	0.761
Overall growth b_3	8.71	[6.98 to 10.5]	<0.001	8.37	[6.46 to 10.3]	<0.001
Slope effect b_4	−1.72	[−4.08 to 1.07]	0.237	−3.07	[−5.63 to −0.53]	<0.05
LNCaP case						
DPN						
Base level b_1	234	[196 to 272]	<0.001	233	[192 to 273]	<0.001
Offset b_2	−22.7	[−78.5 to 33.9]	0.421	−19.5	[−78.5 to 40.5]	0.540
Overall growth b_3	101	[75.4 to 130]	<0.001	52.8	[22.8 to 81.5]	<0.01
Slope effect b_4	−48.9	[−95.6 to −2.91]	<0.05	−41.0	[−83.2 to 2.99]	0.072
ENL						
Base level b_1	234	[194 to 275]	<0.001	233	[191 to 276]	<0.001
Offset b_2	−8.31	[−72.6 to 53.3]	0.784	−5.19	[−72.1 to 60.9]	0.862
Overall growth b_3	101	[74.5 to 130]	<0.001	52.7	[22.6 to 81.7]	<0.01
Slope effect b_4	−81.1	[−125 to –36.1]	<0.01	−45.1	[−92.2 to 1.02]	0.058
4T1 case						
Doxorubicin						
Overall growth b_3	68.4	[57.6 to 79.3]	<0.001	68.4	[57.6 to 79.3]	<0.001
Slope effect b_4	−16.8	[−32.4 to −1.40]	<0.05	−16.8	[−32.0 to −1.12]	<0.05
Cyclophosphamide						
Overall growth b_3	68.4	[60.5 to 76.5]	<0.001	68.4	[60.2 to 76.4]	<0.001
Slope effect b_4	−66.5	[−78.3 to −54.7]	<0.001	−66.8	[−78.2 to −55.3]	<0.001

NOTE: The highest posterior density (HPD), 95% confidence intervals, and *P* values were estimated using 100,000 MCMC simulations. Negative estimates for the treatment specific terms (b_2, b_4) indicate potential treatment effects. Model terms b_1 and b_2 were set to zero in studies without a designated tumor target size (DMBA, 4T1). The fixed effects presented in this table are visualized in a parallel fashion in Supplementary Fig. S10.

the treatment contribute to the identified between-tumor differences in terms of their growth profiles and PSA levels.

We further used the modeling framework to predict that a significant slope effect ($P < 0.05$ at 0.8 power) in response to the DPN treatment could be obtained when the number of tumors is 19 per group (Supplementary Fig. S5A). Notably, with the noncategorizing model, the same sample size estimate would be 25, showing the benefits of the categorizing model already in the initial power analysis. The power analysis also predicted that significant offset effect will not be obtained within reasonable animal numbers. The precision analysis showed differences in the model types and treatment periods when assessing treatment effects (Supplementary Fig. S5B); in particular, the relative importance of the initial time points for the statistical precision (Supplementary Fig. S5C).

4T1 case

In cases such as 4T1, where there is no evident within-group tumor heterogeneity, the EM algorithm classifies all the tumors into the growing category, and therefore the categorizing and noncategorizing models gave the same results (Table 2). More specifically, after adjustment to quadratic growth using residual plots (Supplementary

Published OnlineFirst June 27, 2012; DOI: 10.1158/1078-0432.CCR-11-3215

Table 3. *Post hoc* association analysis of the detected tumor growth categories

	Treatment classes (% within category)			Histologic classes[a] (% within category)			
DMBA case	**Control**	**Treatment**	***P***	**Poorly differentiated**	**Well differentiated**	**Atrophic**	***P***
ENL low dose							
Growing	4 (44%)	5 (56%)		4 (67%)	2 (33%)	0 (0%)	
Poorly growing	9 (53%)	8 (47%)	1.000	2 (17%)	7 (58%)	3 (25%)	0.156
ENL high dose							
Growing	4 (67%)	2 (33%)		3 (60%)	1 (20%)	1 (20%)	
Poorly growing	9 (45%)	11 (55%)	0.645	2 (13%)	10 (67%)	3 (20%)	0.069
MCF-7 case				**ERβ expression[b] (per 1,000 cells)**			
LAR low dose							
Growing	14 (48%)	15 (52%)		248.1 ± 238.7			
Poorly growing	1 (17%)	5 (83%)	0.207	82.0 ± 56.6			0.115
LAR high dose							
Growing	14 (56%)	11 (44%)		213.0 ± 127.0			
Poorly growing	1 (10%)	9 (90%)	0.022	329.7 ± 32.7			0.008
LNCaP case				**PSA concentration[b] (at sacrifice, μg/L)**			
DPN							
Growing	6 (67%)	3 (33%)		97.3 ± 48.3			
Poorly growing	6 (46%)	7 (54%)	0.415	29.3 ± 17.7			0.005
ENL							
Growing	6 (60%)	4 (40%)		99.1 ± 45.5			
Poorly growing	6 (60%)	4 (40%)	1.000	29.1 ± 15.4			0.001

NOTE: Underlining indicates statistical significance ($P \leq 0.05$).

[a]Some of the tumors could not be histologically typed

[b]Values expressed as mean ± SD

Fig. S6), it was confirmed that doxorubicin resulted in regressed tumor growth profiles ($P < 0.05$), whereas cyclophosphamide completely stabilized the growth of each treated tumor ($P < 10^{-6}$).

To test the relative benefits of the categorizing model in a setting where the underlying growth categories and true treatment effect were predefined, we constructed a simulated dataset by combining the doxorubicin- and cyclophosphamide-treated tumors into a single treatment group. The EM algorithm separated the sources of these growth profiles with 100% accuracy within the control and treatment groups (Fig. 3A). The categorizing model also enabled detection of a significant treatment slope effect ($P = 0.002$), which remained undetected by the noncategorizing model (Fig. 3B). This is due to the inability of the noncategorizing model to adjust for the distinct sources of intertumor variation, leading to poor model fit and multimodality in the slope estimates, which could be corrected by taking into account the tumor heterogeneity with the categorizing model (Fig. 3C).

Finally, we also conducted simulations under the null hypothesis of no true treatment effect (Supplementary Material). As expected, an increase in the type-I error appeared under such situation if the categorization approach was applied to homogeneous data or if the noncategorizing approach was applied to heterogeneous data (Supplementary Table S2). The model diagnostic tools should therefore be used to make informed decisions about the model type and structure that is most preferred for the dataset under analysis.

Discussion

This study showed (i) the benefits of modeling the growing and poorly growing categories in terms of improved statistical inference (e.g., DMBA and 4T1 cases); (ii) how the detected categories may be associated with interesting biologic factors, such as endogenous ERβ levels in the MCF-7 case, which provide insights into the underlying tumor heterogeneity; and (iii) how the framework can provide informed suggestions on designing more effective tumor growth experiments in terms of sufficient sample sizes and most informative treatment periods (LNCaP case). The generic modeling framework can also be extended to include additional covariates, such as quadratic growth profiles (Supplementary Fig. S6) or probabilistic tumor categorization (Supplementary Fig. S2). For instance, as the heterogeneity in the growth profiles in the MCF-7 and LNCaP studies was not so clear-cut, continuous growth

Published OnlineFirst June 27, 2012; DOI: 10.1158/1078-0432.CCR-11-3215

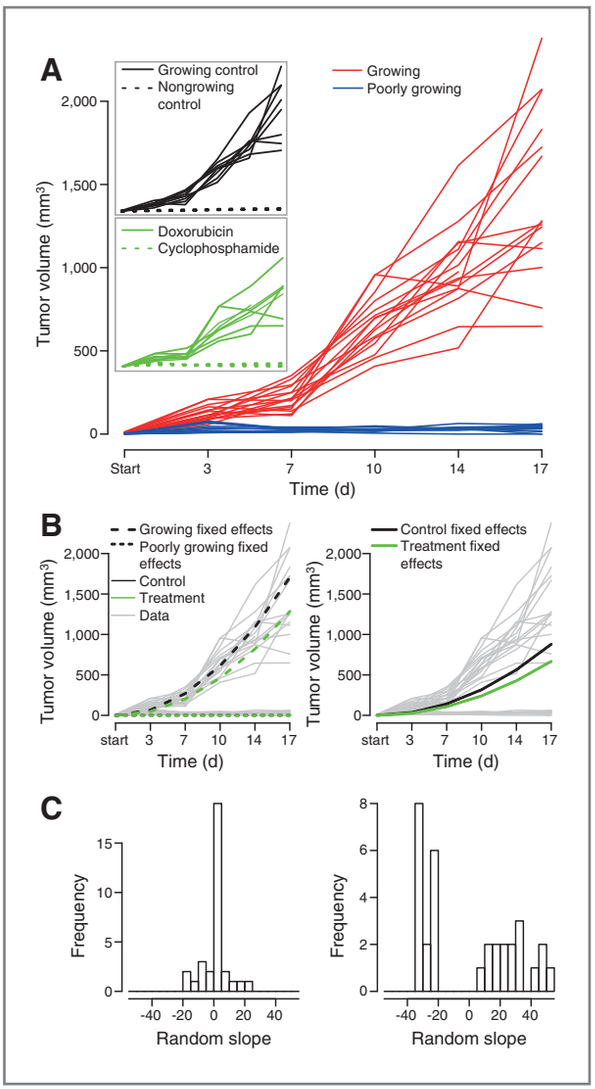

Figure 3. Modeling the simulated 4T1 dataset with and without categorization. A, the EM algorithm and the categorizing model correctly identified the growing and poorly growing subgroups both in the combined control group (original controls and square root of their response traces) and in the combined treatment group (doxorubicin- and cyclophosphamide-treated tumors). B, the categorizing approach detects the doxorubicin-specific treatment effect (left, $P = 0.002$), which is missed by the noncategorizing approach (right, $P = 0.441$). The fixed effect estimates of the noncategorizing model are not feasible due to the assumption of homogeneous growth profiles. C, the random-effects of the categorizing approach show reasonable model fit (left), whereas the random slopes of the noncategorizing approach exhibit severe multimodality (right), suggesting that the growth profiles indeed originate from 2 distinct distributions (i.e., tumor subcategories).

covariates were further used to show that such probabilistic categorization resulted in similar conclusions as obtained from the binary categorization (Supplementary Table S3).

Existing statistical approaches and their limitations

Tumor growth profiles have traditionally been analyzed using univariate statistical approaches that do not fully take into account the tumor heterogeneity within the treatment and control groups. These approaches are typically based on the comparison of tumor sizes at a prespecified time point, using statistical methods such as *t* tests and ANOVA, or their rank-based alternatives such as Wilcoxon–Mann–Whitney and Kruskal–Wallis test (1, 4–12). Another commonly used end point is the time until tumor size doubling, which is analyzed using statistical methods from survival analysis such as the log-rank test (1). There are, however, some potential pitfalls in the use of such single end point approaches. First, an invalid choice of the evaluation time point or the target tumor size may lead to substantial loss of information, in case a large fraction of tumors have not reached the predefined endpoint (32, 33). Second, any single end point is unpowered to detect treatment mechanisms behind dynamic patterns of tumor growth (12). This was exemplified in the DMBA case, where only 2 of the 9 time points showed a significant treatment effect in the original ANOVA-based analysis (6), making the inference upon the efficacy of the dietary intervention more difficult.

Longitudinal statistical modeling methods have also been developed for tumor growth experiments, but these are often restricted to rather specific study designs or questions, and lack effective modeling of intertumor heterogeneity (1). Related approaches that share similar methodologies include, for instance, a standard *t* test together with an EM algorithm as well as Bayesian modeling approaches for testing differences in treatment regimens (13–15). Other authors have developed a nonlinear method for summing 2 exponential functions (16), or a nonparametric approach for estimating tumor growth profiles using penalized spline functions (17). However, even if these models can deal, for instance, with missing and censored data values, other important characteristics of the growth profiles, such as tumor regression or growth rates, cannot be estimated using such approaches. Finally, many of the more advanced statistical models introduced for analyzing tumor growth experiments are not implemented as user-friendly software packages, which hinders their routine use in data analysis.

Recently, an interesting Bayesian hierarchical change-point (BHC) model was proposed for analyzing long treatment experiments (12). The model assumes that the treated tumors will first suppress in response to the treatment, then reach a minimum, and later, rebound with both the decline and the regrowth curves assumed being linear on the log scale. The main difference between our framework and the BHC models is that the latter categorizes the growth profile of each individual tumor into these specific growth periods (i.e., it models intratumor variability), whereas our model categorizes the given set of tumor profiles into growing and poorly growing classes (i.e., it models intertumor variability). The BHC model is especially useful for estimating regression period and nadir tumor volume for such tumors that contain measurements below the limit of quantitation leading to missing values and censored data (1). This is often the case when assessing more aggressive treatment options, which can totally regress the tumor growth and the main focus lies on testing rebound effects and possible

Published OnlineFirst June 27, 2012; DOI: 10.1158/1078-0432.CCR-11-3215

side effects. On the other hand, when experimenting with less aggressive treatment alternatives, more subtle treatment effects are easily missed in case the intertumor heterogeneity is not properly taken into account.

Benefits of the categorizing mixed-effects model

To our knowledge, there are no existing approaches towards modeling the growing and poorly growing tumor categories, even if the presence of such categories in the tumor growth experiments has been long evident (2–8). While there are various approaches to reduce model fit heteroscedasticity, such as the Box-Cox or logarithmic transformations, these cannot model the intrinsic heterogeneity encountered within control and treatment groups. This study showed that when the observed within- and between-group variation is effectively modeled, it is possible to improve the sensitivity of the treatment evaluation through relevant model parameters. In particular, the slope parameter was shown informative when evaluating the efficacy of dietary plant lignans. Our modeling framework also enables comparison of different experimental designs in terms of their associated study power, precision, and sample size estimates, something that is rarely available from other modeling works. However, it should be appreciated that the operation of these modeling tools depends on the data under analysis. Therefore, data visualization and model diagnostics should always be used to confirm that the model assumptions are fulfilled and the model results are valid (see Supplementary Methods for details).

To promote its widespread application in tumor growth studies, we have made publicly available the modeling framework in the form of an R package, named XenoCat, with implementation, source code, user-instructions, and step-by-step example available (28). In contrast to most existing models, our framework can be robustly applied to various tumor growth experiments without making strong assumptions about the type or amount of data under analysis. For instance, the 4 case studies analyzed here were conducted using different tumor models, representing a wide range of experimental setups, such as different number of tumors and various response readouts and their distributional characteristics, which can drastically affect the performance of the traditional statistical methods. The model can deal with short or even outlier profiles, which may be present in the data due to various filtering criteria or very aggressively growing tumors, respectively, and which are frequently excluded from the standard statistical analysis. Therefore, the model can use the full information captured in the entire longitudinal profiles to maximize the output of the tumor growth studies.

The novel tumor categorizing algorithm does not only enable calculating interesting growth parameters, but it also allows for detection of hidden subgroups of differentially growing tumors within treatment and control groups that may associate with the underlying tumor biology. In particular, differences observed in the ERβ expression between the growing and poorly growing categories in the MCF-7 breast cancer model are highly intriguing. Previous studies on genetically modified breast cancer cell lines with high constitutive or inducible expression of ERβ show that tumor growth is significantly reduced when the transgene is turned on (31). Our study is the first, to our knowledge, to show the inverse association between tumor growth and endogenous ERβ expression and suggests that endogenous ERβ levels may be regulated by interventions (here, dietary lignans). This phenomenon may be linked to underlying differences in tumor progression mechanisms (34) and can even give insights into treatment resistance (35). Besides providing additional explanations for the detected tumor growth categories, biologic correlates behind the model-captured tumor heterogeneity could thus open up new possibilities for identifying novel targets and treatment opportunities for cancer.

Limitations of the model and its future extensions

A number of simplifying assumptions were made here to make the implemented model as robust and flexible as possible. The methodology proposed here is based on linear mixed-effects models with dichotomous categorization and assumption that the poorly growing profiles are approximately horizontal. However, in cases where deemed appropriate, the generic model can be extended to more complex settings, including nonlinear growth patterns or several growth categories with non-zero slope parameters or probabilistic tumor categorization, allowing, for instance, partially overlapping groups such as growing, regressing, and stabilizing profiles (6). Another interesting future question we intend to tackle is that whether combining multiple phenotypic readouts for treatment response, such as tumor sizes and PSA levels, would improve statistical power in the case of the prostate cancer model. The current implementation of the power analysis also assumes complete data, but missing values, either informatively censored or missing-at-random (36), could be incorporated in the future work. Finally, the computationally, rather intensive, power calculations could easily be split into parallel processes for maximal computational efficiency (Supplementary Fig. S7).

Disclosure of Potential Conflicts of Interest

M.I. Suominen and E. Alhoniemi were the employees of Pharmatest Services Ltd (at the time of the study), a company that sells services for preclinical cancer treatment testing. M.I. Suominen is also a shareholder of Pharmatest Services Ltd and has ownership interest (including patents) in the same. No potential conflicts of interest were disclosed by other authors.

Authors' Contributions

Conception and design: T.D. Laajala, J. Corander, S.I. Mäkelä, M. Poutanen, T. Aittokallio

Development of methodology: T.D. Laajala, J. Corander, T. Aittokallio

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): N.M. Saarinen, S. Savolainen, M.I. Suominen, S.I. Mäkelä, M. Poutanen

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): T.D. Laajala, M. Poutanen, T. Aittokallio

Writing, review, and/or revision of the manuscript: T.D. Laajala, J. Corander, S. Savolainen, M.I. Suominen, E. Alhoniemi, S.I. Mäkelä, M. Poutanen, T. Aittokallio

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): T.D. Laajala, N.M. Saarinen, K. Mäkelä, E. Alhoniemi, M. Poutanen

Study supervision: J. Corander, M. Poutanen, T. Aittokallio

Published OnlineFirst June 27, 2012; DOI: 10.1158/1078-0432.CCR-11-3215

Interpretation of the biological relevance of the obtained results: N.M. Saarinen

Acknowledgments

The authors thank Dr. Emmy Verschuren (FIMM) for critical reading and useful suggestions for the manuscript and Pharmatest Services Ltd for providing the data from the 4T1 syngeneic mammary cancer model.

Grant Support

This work was supported by the Finnish Funding Agency for Technology and Innovation (Tekes grant 40106/10 to M. Poutanen and T. Aittokallio and Tekes grant 40226/08 to S. Mäkelä); the Academy of Finland (grants 120569, 133227, and 140880 to T. Aittokallio, grant 251170 to J. Corander, grant 138561 to M. Poutanen, grant 123382 to S. Mäkelä; and grant 115459 to N. M. Saarinen); and the European Research Council (ERC grant 239784 to J. Corander).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received December 13, 2011; revised May 25, 2012; accepted June 8, 2012; published OnlineFirst June 27, 2012.

References

1. Heitjan DF. Biology, models, and the analysis of tumor xenograft experiments. Clin Cancer Res 2011;17:949–52.
2. Enmon R, Yang WH, Ballangrud AM, Solit DB, Heller G, Rosen N, et al. Combination treatment with 17-N-allylamino-17-demethoxy geldanamycin and acute irradiation produces supra-additive growth suppression in human prostate carcinoma spheroids. Cancer Res 2003;63:8393–9.
3. Bedogni B, Welford SM, Kwan AC, Ranger-Moore J, Saboda K, Broome Powell M. Inhibition of phosphatidylinositol-3-kinase and mitogen-activated protein kinase 1/2 prevents melanoma development and promotes melanoma regression in the transgenic TPRas mouse model. Mol Cancer Ther 2006;5:3071–7.
4. Gutman M, Couillard S, Labrie F, Candas B, Labrie C. Effects of the antiestrogen EM-800 (SCH 57050) and cyclophosphamide alone and in combination on growth of human ZR-75-1 breast cancer xenografts in nude mice. Cancer Res 1999;59:5176–80.
5. Saarinen NM, Power K, Chen J, Thompson LU. Flaxseed attenuates the tumor growth stimulating effect of soy protein in ovariectomized athymic mice with MCF-7 human breast cancer xenografts. Int J Cancer 2006;119:925–31.
6. Saarinen NM, Huovinen R, Wärri A, Mäkelä SI, Valentín-Blasini L, Sjöholm R, et al. Enterolactone inhibits the growth of 7,12-dimethylbenz(a)anthracene-induced mammary carcinomas in the rat. Mol Cancer Ther 2002;1:869–76.
7. Galaup A, Opolon P, Bouquet C, Li H, Opolon D, Bissery MC, et al. Combined effects of docetaxel and angiostatin gene therapy in prostate tumor model. Mol Ther 2003;7:731–40.
8. Ribonson SP, Jordan VC. Reversal of the antitumor effects of tamoxifen by progesterone in the 7,12-dimethylbenzanthracene-induced rat mammary carcinoma model. Cancer Res 1987;47:5386–90.
9. Terada N, Shimizu Y, Kamba T, Inoue T, Maeno A, Kobayashi T, et al. Identification of EP4 as a potential target for the treatment of castration-resistant prostate cancer using a novel xenograft model. Cancer Res 2010;70:1606–15.
10. Takahara K, Tearle H, Ghaffari M, Gleave ME, Pollak M, Cox ME. Human prostate cancer xenografts in lit/lit mice exhibit reduced growth and androgen-independent progression. Prostate 2011;71:525–37.
11. Shusterman S, Grupp SA, Barr R, Carpentieri D, Zhao H, Maris JM. Angiogenesis inhibitor TNP-470 effectively inhibits human neuroblastoma xenograft growth, especially in the setting of subclinical disease. Clin Cancer Res 2001;7:977–84.
12. 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.
13. Tan M, Fang HB, Tian GL, Houghton PJ. Small sample inference for incomplete longitudinal data with truncation and censoring in tumour xenograft models. Biometrics 2002;58:612–20.
14. Fang HB, Tian GL, Tan M. Hierarchical models for tumour xenograft experiments in drug development. J Biopharm Stat 2004;14:931–45.
15. Tan M, Fang HB, Tian GL, Houghton PJ. Repeated-measures models with constrained parameters for incomplete data in tumour xenograft experiments. Stat Med 2005;24:109–19.
16. Liang H, Sha N. Modeling antitumor activity by using a non-linear mixed-effects model. Math Biosci 2004;189:61–73.
17. Liang H. Modeling antitumor activities in xenograft tumor treatment. Biometrical J 2005;3:358–68.
18. Verbeke G, Lesaffre E. A linear mixed-effects model with heterogeneity in the random-effects population. J Am Stat Assoc 1996;91:217–21.
19. Saarinen NM, Wärri A, Dings RP, Airio M, Smeds AI, Mäkelä S. Dietary lariciresinol attenuates mammary tumor growth and reduces blood vessel density in human MCF-7 breast cancer xenografts and carcinogen-induced mammary tumors in rats. Int J Cancer 2008;123:1196–204.
20. Reeves PG, Nielsen FH, Fahey GC Jr. AIN-93 purified diets for laboratory rodents: final report of the American Institute of Nutrition ad hoc writing committee on the reformulation of the AIN-76A rodent diet. J Nutr 1993;123:1939–51.
21. Suominen MI, Käkönen R, Käkönen SM, Halleen JM. Diverging effects of doxorubicin, paclitaxel and cyclophosphamide on 4T1 mouse breast cancer primary tumor and metastases. Poster at joint AACR-MRS meeting: Metastasis and the tumor microenvironment, September 12–15 2010, Philadelphia, USA. Available from: www.pharmatest.fi
22. Du G, Lin H, Wang M, Zhang S, Wu X, Lu L, et al. Quercetin greatly improved therapeutic index of doxorubicin against 4T1 breast cancer by its opposing effects on HIF-1α in tumor and normal cells. Cancer Chemother Pharmacol 2010;65:277–87.
23. Viola RJ, Provenzale JM, Li F, Li CY, Yuan H, Tashjian J, et al. *In vivo* bioluminescence imaging monitoring of hypoxia-inducible factor 1alpha, a promoter that protects cells, in response to chemotherapy. Am J Roentgenol 2008;191:1779–84.
24. Kamb A. What's wrong with our cancer models? Nat Rev Drug Discov 2005;4:161–5.
25. Bates DM, Maechler M, Bolker B. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-39; 2011 [cited 2012 May 25]. Available from: http://CRAN.R-project.org/package=lme4
26. R Development Core Team. R: A language and environment for statistical computing. R statistical software version 2.14; 2011 [cited 2012 May 25]. Available from: http://www.R-project.org
27. Baayen RH. Modeling data with fixed and random effects. In: Analyzing linguistic data, a practical introduction to statistics using R, Cambridge University Press; 2008. p. 242–58.
28. XenoCat-project, R package version 1.0.3; 2012 [cited 2012 May 25]. Available from: http://code.google.com/p/r-xenocat/.
29. Gelman A, Hill J. Multilevel power calculation using fake-data simulation. In:Alvarez RM, Beck NL, Wu LL editors. Data analysis using regression and multilevel/hierarchical models. New York: Cambridge University Press; 2007. p. 449–54.
30. Stroup WW. Mixed model procedures to assess power, precision and sample size in the design of experiments. In: ASA Proceedings of the biopharmaceutical section. Alexandria, VA: American Statistical Association; 1999. p. 15–24.
31. Hartman J, Lindberg K, Morani A, Inzunza J, Ström A, Gustafsson JA. Estrogen receptor beta inhibits angiogenesis and growth of T47D breast cancer xenografts. Cancer Res 2006;66:11207–13.

Published OnlineFirst June 27, 2012; DOI: 10.1158/1078-0432.CCR-11-3215

32. Begg AC. Analysis of growth delay data: potential pitfalls. Br J Cancer Suppl 1980;4:93–97.

33. Heitjan DF, Manni A, Santen RJ. Statistical analysis of *in vivo* tumor growth experiments. Cancer Res 1993;53:6042–50.

34. Horimoto Y, Hartman J, Millour J, Pollock S, Olmos Y, Ho KK, et al. ERβ1 represses FOXM1 expression through targeting ERα to control cell proliferation in breast cancer. Am J Pathol 2011;179:1148–56.

35. Hopp TA, Weiss HL, Parra IS, Cui Y, Osborne CK, Fuqua SA. Low levels of estrogen receptor β protein predict resistance to Tamoxifen therapy in breast cancer. Clin Cancer Res 2004;10:7490–9.

36. Aittokallio T. Dealing with missing values in large-scale studies: microarray data imputation and beyond. Brief Bioinform 2010;11:253–64.

Published OnlineFirst June 27, 2012; DOI: 10.1158/1078-0432.CCR-11-3215

Clinical Cancer Research

Improved Statistical Modeling of Tumor Growth and Treatment Effect in Preclinical Animal Studies with Highly Heterogeneous Responses *In Vivo*

Teemu D. Laajala, Jukka Corander, Niina M. Saarinen, et al.

Clin Cancer Res 2012;18:4385-4396. Published OnlineFirst June 27, 2012.