Bayesian statistics for longitudinal studies in biomedical research

Their application and use in biomedical research

Ariel Mundo[*] Timothy J. Muldoon[†] John R. Tipton[‡]

Paper outline

The paper Introduction has been updated, proposed sections appear at the end of the document as well as an initial graph.

Background

A longitudinal study is defined as that which is designed to repeatedly measure a variable of interest in a group (or groups) of subjects. In biomedical research, this type of study arises when the investigator intends to observe the evolution of the effect of a certain treatment across time, rather than analyzing it at a single time point (a cross-sectional study). Clinical examples of this approach in biomedical research include studies on breast and neck cancer(Sio et al. 2016; Kamstra et al. 2015); in the first case, weekly measurements of skin toxicities in patients with radiation-induced dermatitis were taken for up to 8 weeks; whereas in the latter mouth opening was assessed at 6,12, 18, 24 and 36 months after radiotherapy (RT). Longitudinal studies have used also to measure tumor response (Roblyer et al. 2011; Tank et al. 2020; Pavlov et al. 2018; Demidov et al. 2018), antibody expression(Ritter et al. 2001; Roth et al. 2017), and cell metabolism(Jones et al. 2018; Skala et al. 2010). From a statistical standpoint, a longitudinal study presents advantages over a cross-sectional approach:it requires a lower number of subjects to reach a certain statistical power, and besides it being able to track the previously mentioned time-effect evolution on a group-by-group basis, it allows to determine the variability of the response within subjects (Guo et al. 2013; Fitzmaurice, Laird, and Ware 2012). In other words,a longitudinal study permits to quantify how the variable changes within each subject across time.

Traditionally, a "frequentist" approach is used in biomedical research to derive inferences from the results of a longitudinal study. Such statistical view derives its name from the fact that it regards probability as a limiting frequency [wagenmakers2008] and its application is based on a null hypothesis test using the *analysis of variance over repeated measures* (repeated measures ANOVA or rm-ANOVA). This methodology makes two key assumptions regarding longitudinal data: a constant correlation exists across same-subject measurements, and observations from each subject are obtained at all time points through the study (Schober and Vetter 2018; Gueorguieva and Krystal 2004).

However, constant correlation is frequently unjustified as its value tends to diminish between measures when the time interval between them increases(Ugrinowitsch, Fellingham, and Ricard 2004), and the violation of this assumption increases the false positivity rate (Lane 2016). Moreover, it is unlikely that complete observations in all subjects are obtained in a biomedical study due to reasons that can be specific to different situations. In a clinical trial, voluntary withdrawal from one or multiple patients can occur, whereas attrition

[*]Department of Biomedical Engineering, University of Arkansas, Fayetteville
[†]Department of Biomedical Engineering, University of Arkansas, Fayetteville
[‡]Department of Mathematical Sciences, University of Arkansas, Fayetteville

in animals due to injury or weight loss can occur in preclinical experiments, and in both cases it is possible that unexpected complications with equipment or supplies arise, preventing the researcher from collecting measurements at a certain time point.

When the aforementioned issues arise, rm-ANOVA requires to exclude all subjects with missing observations from the analysis, thereby increasing costs for the study if the desired statistical power is not met with the remaining observations as it makes necessary to enroll more subjects; and raising the possibility that the rejection of those partial observations limits the demonstration of significant differences between groups. Additionally, rm-ANOVA uses a *post hoc* analysis to assess the relevance between the measured response in different groups. This analysis is based on multiple repeated comparisons to estimate a *p-value*, a metric that is widely used as a measure of significance. Because the p-value is highly variable (Halsey et al. 2015; Nuzzo 2014), multiple comparisons can inflate the false positivity rate (Liu, Cripe, and Kim 2010),consequently biasing the conclusions of the study.

During the last decade, the biomedical community has started to recognize the limitations of a rm-ANOVA approach in the analysis of longitudinal information. This is exemplified by the pioneering use of linear mixed effects models (LMEMs) in certain groups to analyze tumor longitudinal data (Skala et al. 2010; Vishwanath et al. 2009). Briefly, these models incorporate *fixed effects*, which correspond to the levels of experimental factors in the study (e.g. the different drug regimens in a clinical trial), and *random effects*, which account for random variation within the population (Pinheiro and Bates 2006).These models are more flexible than rm-ANOVA as they can accommodate missing observations for multiple subjects, and allow different modeling strategies for the variability within each measure in every subject (West, Welch, and Galecki 2014; Pinheiro and Bates 2006).On the other hand, they impose restrictions in the distribution of the errors of the model and random effects (Gueorguieva and Krystal 2004; Schielzeth et al. 2020).

Additionally, another assumption is made in both rm-ANOVA and LMEMs models, where a linear relationship is expected between the observed response and the covariates across the study (Pinheiro and Bates 2006).This common assumption to both rm-ANOVA and LMEMs causes the models to be restrictive in their inferences when used in longitudinal data that does not follow a linear trend. In biomedical research, this particular behavior in longitudinal has been reported in studies of tumor response to radio/chemotherapy in preclinical and clinical settings(Vishwanath et al. 2009; Roblyer et al. 2011; Tank et al. 2020; Skala et al. 2010; Demidov et al. 2018), and wound healing and metabolism(Jones et al. 2018; Grice et al. 2010; Young and Grinnell 1994). This studies have shown that that the collected signal does not follow a linear trend over time, and presents high variability at different time points, making the estimations of a LMEM or rm-ANOVA model inconsistent with the pattern of the observed variations. Additionally, although it is possible that a post hoc analysis is able to find "significance"(*p-value*<0.05) by using multiple comparisons between the model terms, this estimator would be inherently biased because of the lack of fit between the information and the model.

As the "frequentist" rm-ANOVA and the more advanced LMEM approach are both limited in the analysis of non-linear longitudinal information, there is a need for biomedical researchers to explore the use of additional statistical tools that allow the information (and not an assumed distribution) to determine the fit of the model while enabling inferences that are both adequate and consistent from a statistical perspective.

Generalized additive models (GAMs) are a subset of generalized linear models that use *smooth functions* (hencefort *smooths*) to estimate the parameters of a model. They have been used in palaeolimnology, ecology and clinical studies to model longitudinal data(Woolway et al. 2016; Hefley et al. 2017; Ko et al. 2007). Briefly, GAMs use a combination of multiple functions (basis functions)to construct the smooths of the model (Wood 2017). Their main advantage over LMEMs and rm-ANOVA is that the model specification is directed by the *smooths* rather than by a parametric relationship. This allows a consistent fit of the model with the data, and estimations of significance using the terms of the model. However, certain assumptions about the data are necessary: a normal distribution and constant variance of the residuals with the mean response. Therefore, GAMs provide a more suitable statistical method to analyze biomedical longitudinal data, when these assumptions of the model are met by the data.

However, it is possible that the assumptions of GAMs for longitudinal data do not hold under certain circumstances. In that case, the field of *Bayesian statistics* represents a relatively new area of Statistics that does not rely on *p-values* and hypothesis tests to analyze information. Bayesian statistics can work with missing observations, allow the data (and not an underlying assumed distribution) to determine the outcome in regard to significance and are able to expand the number of comparisons and inferences derived form the analysis. On the other hand, the shift that Bayesian statistics represent from the traditional "frequentist" statistical view in research, the computational tools required for its implementation, and the underlying mathematical theory have limited the use of this approach in the biomedical research community. However,Bayesian theory is intuitive and shares some principles with "frequentist" statistics, and there is an increasing use and recognition of the advantages of their use across different areas of biomedical research such as clinical trial design and imaging(Biswas et al. 2009; Kelter 2020; Kwon et al. 2020; Zhou 2017).

Additionally,the current development in computational tools, specifically the programming language R, enable a rapid implementation of GAMs and Bayesian models for longitudinal data.In particular, R has an extensive collection of documentation, functions and libraries that speed up the initial stages of the analsys and that enable the use of complex statistical methods without requiring a specialized set of programming skills from the user.

Therefore, this study focuses in three areas in the analysis of longitudinal data from a biomedical perspective. First, it presents the limitations of (rm-ANOVA) and LMEMs over longitudinal data, and explains how these limitations in turn affect the results of the analysis. Secondly, it uses R to simulate non-linear longitudinal data following previously reported values in the literature, and presents the implementation of GAMs as a statistical tool for longitudinal data. And finally, it introduces Bayesian statistics and presents their implementation with GAMs to demonstrate the differences and benefits of this approach. with an emphasis in reproducibility by providing the code, simulated dataset and a step-by-step guide of the computational implementation of different models, this study aims to encourage the exploration of modern statistical methods for biomedical longitudinal data, and to improve the statistical standards in biomedical research.

- Why LMEMs are better than ANOVA
- How LMEMs (using splines) and a Bayesian analysis can be used to analyze longitudinal data

https://www.frontiersin.org/articles/10.3389/fevo.2018.00149/full

Section 1:

Challenges presented by longitudinal studies:

1 The "frequentist" case for longitudinal data

The *repeated measures analysis of variance* (rm-ANOVA) is the standard for the statistical analysis of longitudinal data, and there are key assumptions that are made in order to make the model valid. From a practical view, they can be divided in three areas: linear relationship between covariates and response, constant correlation between measurements, and complete observations for all subjects. Each one of these assumptions is discussed below.

1.1 Linear relationship In a biomedical longitudinal study, two or more groups of subjects (humans, mice, samples) are subject to a different treatments (e.g. group of mice receiving a novel drug vs. a group that receives a placebo), and measurements from each subject within each group are collected at specific time points. Moreover, it is assumed that the collected response has two components: a *fixed* and a *random* component. The *fixed* component can be understood as a constant value in the response which the researcher intends to measure, i.e, the effect of a novel drug in a subject.The *random* component can be defined as "noise" caused by some some factors that are not of interest to the researcher, i.e., if the concentration of a drug is measured in some subjects within the same group in the early hours of the morning while others are measured in the afternoon, the researcher might consider this variability in the collection time of the

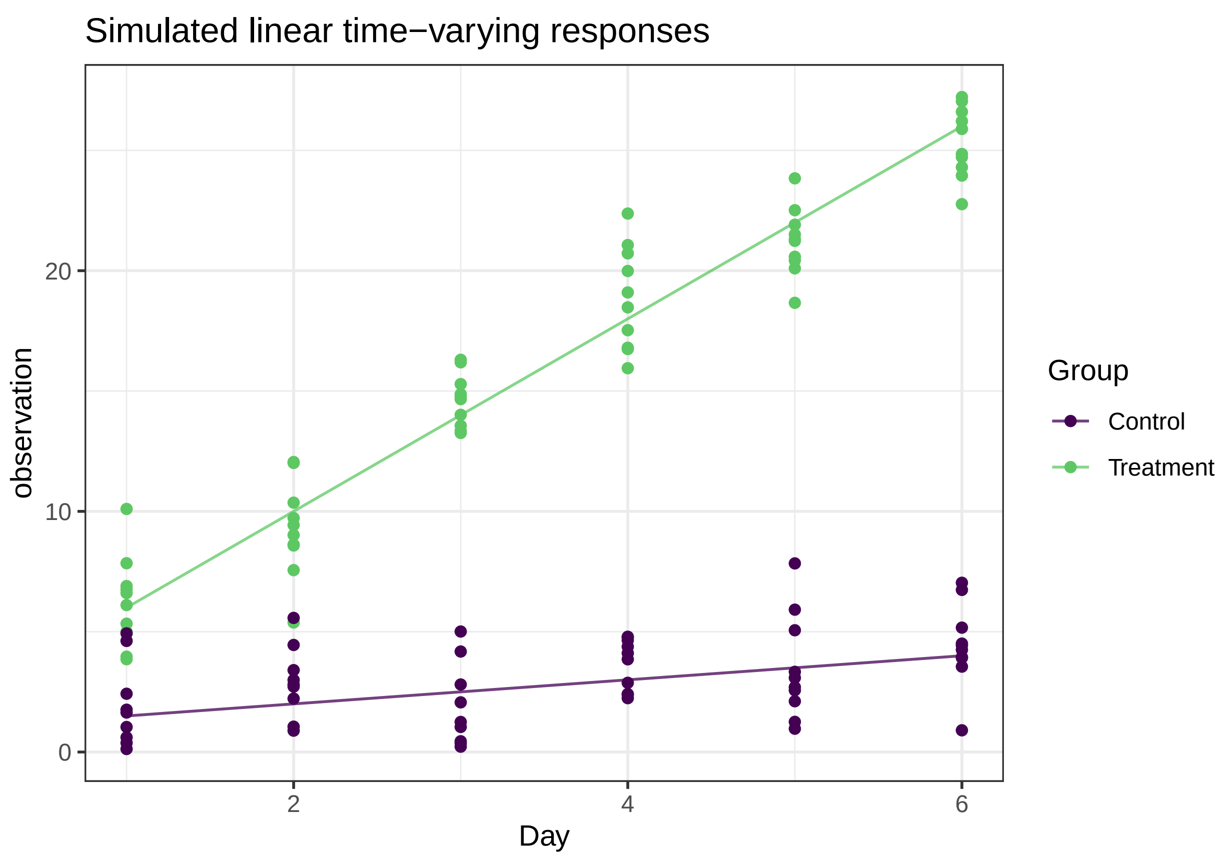

Figure 1: Simulated longitudinal data with a linear trend.

measurement to introduce some "noise" in the signal. As their name suggests, this "random" variabliity needs to be modeled as a variable rather than as a constant value.

Mathematically speaking, if a normally distributed response y is measured repeatedly at t time points from subjects in p groups, where each group has a certain n_p number of subjects, the the model for the response y_{hij} becomes:

$$y_{hij} = \mu + \gamma_h + \tau_j + (\gamma\tau)_{hj} + \pi_{i(h)} + e_{hij}(\#eq:labelANOVA) \tag{1}$$

Where

$i = 1, ..., n_h$ $j = 1, ..., t$, $h = 1, ..., p$; with $\pi_i h \sim N(0, \sigma_\epsilon^2)$ (independently normally distributed) and $e_{hij} \sim N(0, \sigma_\epsilon^2)$

In this model, μ represents the group mean, γ_h is the *fixed effect* of group h, τ_j is the fixed effect of time j, and $(\gamma\tau)_{hj}$ represents the interaction of time and group effects. The term π_{ij} represents the *random effects* for each subject within each group.Finally, e_{hij} represents the independent random error terms, which need to be normally distributed with mean 0 (Davis 2002). The model then, is a linear combination of terms, and if plotted, it would a straight line.

Question: How can one make a plot that tests the "limits" of the model, i.e how much "wiggliness" can this model accomodate? Make a plot to show the behavior of the model

1.2 Covariance and correlation In a longitudinal study, the fact that multiple measures are taken on the same subject creates a a *covariance* issue that needs to be incorporated into the model. In this case, *covariance* can be defined as the dependency (linear?) between two different values, e.g. if higher values of a variable correspond to higher values of another value, the covariance is positive in this case. Because the value for the covariance is unnormalized, the *correlation* is used (Weiss 2005) which is normalized. While it is not immediately apparent the reason for the need to specify the covariance, the main reason is that computationally, the model from In a longitudinal study, the measured response can show different profiles for each subject. For some subjects, there is little variability in the response across time, while for others

make plot that explains between and within correlation

In this case, *correlation* can be defined as the relationship of the effect between different variables, e.g. how much one variable changes in regard to the other. This concept can be further decomposed in *between-subject* correlation and *within-subject correlation.* In the first case, *between-subject correlation* measures the correlation between subject means, whereas *within-subject correlation* describes how much a measurement in a subject changes in regard to another to another measurement *in the same subject* (Roy 2006). between repeated measures on the same subject can be defined as the relationship between the variability . The *correlation* of the measures can be practically defined as is defined as the relationship between
Missing observations, and correlation between measurements. How rm-ANOVA is limited by missing observations, and how both rm-ANOVA and LMEMs are limited with data that does not follow a linear trend (equations for both situations and the fit they produce.) #### 1.1 Constant correlation

While the *repeated measures analysis of variance* (rm-ANOVA) is the norm in the biomedical community to analyze longitudinal data, two frequent situations that arise in longitudinal studies limit its applicability. Missing observations, and correlation between measurements. How rm-ANOVA is limited by missing observations, and how both rm-ANOVA and LMEMs are limited with data that does not follow a linear trend (equations for both situations and the fit they produce.)

One of the key assumptions for rm-ANOVA is the constant correlation among measurements. This frequently not the case as the correlation reduces as the time interval between two measurements increases (Ugrinowitsch, Fellingham, and Ricard 2004). From a practical

1.2 the case of GMEMs using splines and how they work and how they are better than rm-ANOVA

1.3 Bayesian brief introduction, and compare the results of 1.2 to the results of Bayesian

- Section 2: Implementation of both LMEMs and Bayesian and their results

Present the implementation of a spline-fitted model in R, using data simulated from (Vishwanath et al. 2009)

References

Biswas, Swati, Diane D Liu, J Jack Lee, and Donald A Berry. 2009. "Bayesian Clinical Trials at the University of Texas Md Anderson Cancer Center." *Clinical Trials* 6 (3): 205–16.

Davis, Charles S. 2002. *Statistical Methods for the Analysis of Repeated Measurements*. Springer Science & Business Media.

Demidov, Valentin, Azusa Maeda, Mitsuro Sugita, Victoria Madge, Siddharth Sadanand, Costel Flueraru, and I Alex Vitkin. 2018. "Preclinical Longitudinal Imaging of Tumor Microvascular Radiobiological Response with Functional Optical Coherence Tomography." *Scientific Reports* 8 (1): 1–12.

Fitzmaurice, Garrett M, Nan M Laird, and James H Ware. 2012. *Applied Longitudinal Analysis*. Vol. 998. John Wiley & Sons.

Grice, Elizabeth A, Evan S Snitkin, Laura J Yockey, Dustin M Bermudez, Kenneth W Liechty, Julia A Segre, NISC Comparative Sequencing Program, and others. 2010. "Longitudinal Shift in Diabetic Wound Microbiota Correlates with Prolonged Skin Defense Response." *Proceedings of the National Academy of Sciences* 107 (33): 14799–14804.

Gueorguieva, Ralitza, and John H Krystal. 2004. "Move over Anova: Progress in Analyzing Repeated-Measures Data Andits Reflection in Papers Published in the Archives of General Psychiatry." *Archives of General Psychiatry* 61 (3): 310–17.

Guo, Yi, Henrietta L Logan, Deborah H Glueck, and Keith E Muller. 2013. "Selecting a Sample Size for Studies with Repeated Measures." *BMC Medical Research Methodology* 13 (1): 100.

Halsey, Lewis G, Douglas Curran-Everett, Sarah L Vowler, and Gordon B Drummond. 2015. "The Fickle P Value Generates Irreproducible Results." *Nature Methods* 12 (3): 179–85.

Hefley, Trevor J, Kristin M Broms, Brian M Brost, Frances E Buderman, Shannon L Kay, Henry R Scharf, John R Tipton, Perry J Williams, and Mevin B Hooten. 2017. "The Basis Function Approach for Modeling Autocorrelation in Ecological Data." *Ecology* 98 (3): 632–46.

Jones, Jake D, Hallie E Ramser, Alan E Woessner, and Kyle P Quinn. 2018. "In Vivo Multiphoton Microscopy Detects Longitudinal Metabolic Changes Associated with Delayed Skin Wound Healing." *Communications Biology* 1 (1): 1–8.

Kamstra, JI, PU Dijkstra, M Van Leeuwen, JLN Roodenburg, and JA Langendijk. 2015. "Mouth Opening in Patients Irradiated for Head and Neck Cancer: A Prospective Repeated Measures Study." *Oral Oncology* 51 (5): 548–55.

Kelter, Riko. 2020. "Bayesian Alternatives to Null Hypothesis Significance Testing in Biomedical Research: A Non-Technical Introduction to Bayesian Inference with Jasp." *BMC Medical Research Methodology* 20: 1–12.

Ko, Fanny WS, Wilson Tam, Tze Wai Wong, Doris PS Chan, Alvin H Tung, Christopher KW Lai, and David SC Hui. 2007. "Temporal Relationship Between Air Pollutants and Hospital Admissions for Chronic Obstructive Pulmonary Disease in Hong Kong." *Thorax* 62 (9): 780–85.

Kwon, Yongchan, Joong-Ho Won, Beom Joon Kim, and Myunghee Cho Paik. 2020. "Uncertainty Quantification Using Bayesian Neural Networks in Classification: Application to Biomedical Image Segmentation." *Computational Statistics & Data Analysis* 142: 106816.

Lane, D. 2016. "The Assumption of Sphericity in Repeated-Measures Designs: What It Means and What to Do When It Is Violated." *Quantitative Methods for Psychology* 12: 114–22.

Liu, Chunyan, Timothy P Cripe, and Mi-Ok Kim. 2010. "Statistical Issues in Longitudinal Data Analysis for Treatment Efficacy Studies in the Biomedical Sciences." *Molecular Therapy* 18 (9): 1724–30.

Nuzzo, Regina. 2014. "Scientific Method: Statistical Errors." *Nature News* 506 (7487): 150.

Pavlov, Mikhail V, Tatiana I Kalganova, Yekaterina S Lyubimtseva, Vladimir I Plekhanov, German Yurievich Golubyatnikov, Olga Y Ilyinskaya, Anna G Orlova, et al. 2018. "Multimodal Approach in Assessment of the Response of Breast Cancer to Neoadjuvant Chemotherapy." *Journal of Biomedical Optics* 23 (9): 091410.

Pinheiro, José, and Douglas Bates. 2006. *Mixed-Effects Models in S and S-Plus*. Springer Science & Business Media.

Ritter, Gerd, Leonard S Cohen, Clarence Williams, Elizabeth C Richards, Lloyd J Old, and Sydney Welt. 2001. "Serological Analysis of Human Anti-Human Antibody Responses in Colon Cancer Patients Treated with Repeated Doses of Humanized Monoclonal Antibody A33." *Cancer Research* 61 (18): 6851–9.

Roblyer, Darren, Shigeto Ueda, Albert Cerussi, Wendy Tanamai, Amanda Durkin, Rita Mehta, David Hsiang, et al. 2011. "Optical Imaging of Breast Cancer Oxyhemoglobin Flare Correlates with Neoadjuvant Chemotherapy Response One Day After Starting Treatment." *Proceedings of the National Academy of Sciences* 108 (35): 14626–31.

Roth, Eli M, Anne C Goldberg, Alberico L Catapano, Albert Torri, George D Yancopoulos, Neil Stahl, Aurélie Brunet, Guillaume Lecorps, and Helen M Colhoun. 2017. "Antidrug Antibodies in Patients Treated with Alirocumab."

Roy, Anuradha. 2006. "Estimating Correlation Coefficient Between Two Variables with Repeated Observations Using Mixed Effects Model." *Biometrical Journal* 48 (2): 286–301.

Schielzeth, Holger, Niels J Dingemanse, Shinichi Nakagawa, David F Westneat, Hassen Allegue, Céline Teplitsky, Denis Réale, Ned A Dochtermann, László Zsolt Garamszegi, and Yimen G Araya-Ajoy. 2020. "Robustness of Linear Mixed-Effects Models to Violations of Distributional Assumptions." *Methods in Ecology and Evolution* 11 (9): 1141–52.

Schober, Patrick, and Thomas R Vetter. 2018. "Repeated Measures Designs and Analysis of Longitudinal Data: If at First You Do Not Succeed—Try, Try Again." *Anesthesia and Analgesia* 127 (2): 569.

Sio, Terence T, Pamela J Atherton, Brandon J Birckhead, David J Schwartz, Jeff A Sloan, Drew K Seisler, James A Martenson, et al. 2016. "Repeated Measures Analyses of Dermatitis Symptom Evolution in Breast Cancer Patients Receiving Radiotherapy in a Phase 3 Randomized Trial of Mometasone Furoate Vs Placebo (N06c4 [Alliance])." *Supportive Care in Cancer* 24 (9): 3847–55.

Skala, Melissa C, Andrew Nicholas Fontanella, Lan Lan, Joseph A Izatt, and Mark W Dewhirst. 2010. "Longitudinal Optical Imaging of Tumor Metabolism and Hemodynamics." *Journal of Biomedical Optics* 15 (1): 011112.

Tank, Anup, Hannah M Peterson, Vivian Pera, Syeda Tabassum, Anais Leproux, Thomas O'Sullivan, Eric Jones, et al. 2020. "Diffuse Optical Spectroscopic Imaging Reveals Distinct Early Breast Tumor Hemodynamic Responses to Metronomic and Maximum Tolerated Dose Regimens." *Breast Cancer Research* 22 (1): 1–10.

Ugrinowitsch, Carlos, Gilbert W Fellingham, and Mark D Ricard. 2004. "Limitations of Ordinary Least Squares Models in Analyzing Repeated Measures Data." *Medicine and Science in Sports and Exercise* 36: 2144–8.

Vishwanath, Karthik, Hong Yuan, William T Barry, Mark W Dewhirst, and Nimmi Ramanujam. 2009. "Using Optical Spectroscopy to Longitudinally Monitor Physiological Changes Within Solid Tumors." *Neoplasia* 11 (9): 889–900.

Weiss, Robert E. 2005. *Modeling Longitudinal Data*. Springer Science & Business Media.

West, Brady T, Kathleen B Welch, and Andrzej T Galecki. 2014. *Linear Mixed Models: A Practical Guide Using Statistical Software*. CRC Press.

Wood, Simon N. 2017. *Generalized Additive Models: An Introduction with R*. CRC press.

Woolway, R Iestyn, Ian D Jones, Stephen C Maberly, Jon R French, David M Livingstone, Donald T Monteith, Gavin L Simpson, et al. 2016. "Diel Surface Temperature Range Scales with Lake Size." *Plos One* 11 (3): e0152466.

Young, Patty K, and Frederick Grinnell. 1994. "Metalloproteinase Activation Cascade After Burn Injury: A Longitudinal Analysis of the Human Wound Environment." *Journal of Investigative Dermatology* 103 (5): 660–64.

Zhou, Tianjian. 2017. "Bayesian Nonparametric Models for Biomedical Data Analysis." *arXiv Preprint arXiv:1710.09890*.