**Bayesian statistics for longitudinal studies in biomedical research**

*Their application and use in biomedical research*

Ariel Mundo[[1]](#footnote-20)

Timothy J. Muldoon[[2]](#footnote-21)

John R. Tipton[[3]](#footnote-22)

## Paper outline

## Background

A longitudinal study is defined as one where a variable of interest is measured repeatedly in a group (or groups) of subjects. In biomedical research, this type of study is preferred when the intention is to observe the evolution of the effect of treatment across time, rather than analyzing the information at a single timepoint (a cross-sectional study). Clinical examples of this approach in biomedical research include studies on breast and neck cancer(Sio et al. [2016](#ref-sio2016); Kamstra et al. [2015](#ref-kamstra2015)); 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 measured 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](#ref-roblyer2011); Tank et al. [2020](#ref-tank2020); Pavlov et al. [2018](#ref-pavlov2018); Demidov et al. [2018](#ref-demidov2018)), antibody expression(Ritter et al. [2001](#ref-ritter2001); Roth et al. [2017](#ref-roth2017)), and cell metabolism(Jones et al. [2018](#ref-jones2018); Skala et al. [2010](#ref-skala2010)). 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](#ref-guo2013); Fitzmaurice, Laird, and Ware [2012](#ref-fitzmaurice2012)). In other words,a longitudinal study permits to quantify how the variable changes within each subject across time.

Researchers have typically employed a *frequentist* approach to analyze longitudinal data, a statistical view that derives its name from the fact that it regards probability as a limiting frequency [wagenmakers2008]. when applied to longitudinal, a frequentist analysis 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 across same-subject measurements and complete observations from each subject (Schober and Vetter [2018](#ref-schober2018); Gueorguieva and Krystal [2004](#ref-gueorguieva2004)). The first condition is frequently unjustified as the correlation between measures diminishes when the time gap between them increases (Ugrinowitsch, Fellingham, and Ricard [2004](#ref-ugrinowitsch2004)), and in that case violating that assumption increases the type I error rate(Lane [2016](#ref-lane2016)). The reasons behind the second assumption being not met in biomedical research are varied: subjects can withdraw during the course of the study, attrition in animals due to injury or weight loss, or complications can arise that prevent the researcher from collecting measurements at a certain timepoint. When these issues arise, rm-ANOVA requires to exclude the subject with missing observations from the analysis, which can lead to increased costs for the study if the desired statistical power is not met with the remaining subjects, and can lead to the exclusion of valuable information.Additionaly, rm-ANOVA uses a *post hoc* analysis to assess significance in the differences between groups. Because a *post hoc* analysis is based in multiple repeated comparisons, it can inflate the false positivity rate(Liu, Cripe, and Kim [2010](#ref-liu2010)).

Recently, *linear mixed effects models* (LMEMs) have been used by certain groups to analyze biomedical longitudinal data (Vishwanath et al. [2009](#ref-vishwanath2009); Skala et al. [2010](#ref-skala2010)). 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](#ref-pinheiro2006)). These models are more flexible than rm-ANOVA as they can accommodate missing observations in the data, and allow different modeling strategies for the covariance(West, Welch, and Galecki [2014](#ref-west2014)). On the other hand, they impose restrictions in the distribution of the errors and of the random effects, and assume linearity between the response and the parameters(Schielzeth et al. [2020](#ref-schielzeth2020); Pinheiro and Bates [2006](#ref-pinheiro2006)).

The assumption of linearity is a key component of LMEMs and rm-ANOVA, and therefore both models restrict the inferences they can extract from longitudinal data when it does not follow a linear trend. In biomedical research, such behavior in longitudinal data does arise and has been exemplified in different studies, including response to radio/chemotherapy in preclinical and clinical settings (Vishwanath et al. [2009](#ref-vishwanath2009); Roblyer et al. [2011](#ref-roblyer2011); Tank et al. [2020](#ref-tank2020); Skala et al. [2010](#ref-skala2010); Demidov et al. [2018](#ref-demidov2018)), and wound healing and metabolism(Jones et al. [2018](#ref-jones2018); Grice et al. [2010](#ref-grice2010); Young and Grinnell [1994](#ref-young1994)). In such situations, an estimated model via rm-ANOVA or LMEMs may not provide a consistent fit with the observed variations in the variables of interest. Moreover, although it is possible to obtain a “significant” *p-value*( *p*<0.05) due to the well-known variability of this metric (Nuzzo [2014](#ref-nuzzo2014); Halsey [2019](#ref-halsey2019)), the model will lack predictive power and this in turn will compromise the extent of the inferences that can be derived from it.

Since both the *frequentist* (rm-ANOVA) and LMEM approaches are limited in the analysis of non-linear longitudinal information, there is a need for biomedical researchers to use statistical tools that allow the data to determine the fit of the model while enabling inferences that are both adequate and consistent from a statistical standpoint.

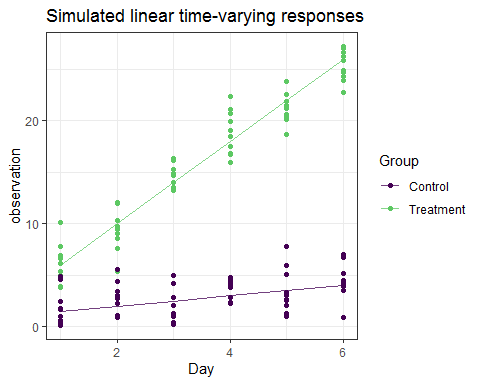
Generalized additive models (GAMs) are a subset of generalized linear models that use *smooth functions* (hencefort *smooths*) to estimate the parameters of a model. They have been used in palaeolimnology, ecology and clinical studies to model longitudinal data(Woolway et al. [2016](#ref-woolway2016); Hefley et al. [2017](#ref-hefley2017); Ko et al. [2007](#ref-ko2007)). Briefly, GAMs use a combination of multiple functions (basis functions)to construct the smooths of the model (Wood [2017](#ref-wood2017)). 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](#ref-biswas2009); Kelter [2020](#ref-kelter2020); Kwon et al. [2020](#ref-kwon2020); Zhou [2017](#ref-zhou2017)). Additionally,the current development in computational tools, specifically the programming language , enable a rapid implementation of Bayesian models for longitudinal data.

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 simulation to generate non-linear longitudinal data and presents the implementation of GAMs as a statistical tool that enables the analysis of data that shows this behavior. And finally, it introduces Bayesian statistics and presents their implementation with GAMs over simulated data to demonstrate the differences and benefits of this approach . With an emphasis on reproducibility by providing the code and dataset used, this study provides biomedical researchers with a clear view of the advantages of different approaches in the statistical analysis of longitudinal data.

* 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>



Simulated longitudinal data with a linear trend.

## Section 1:

### Challenges presented by longitudinal studies:

### 1 The *frequentist* case for longitudinal data

While the *repeated measures analysis of variance* (rm-ANOVA) is the standard for the statistical analysis of longitudinal data, 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 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 is measured repeatedly at time points from subjects in groups, where each group has a certain number of subjects, the the model for the response becomes:

Where

, ; with (independently normally distributed) and

In this model, represents the group mean, is the *fixed effect* of group , is the fixed effect of time , and represents the interaction of time and group effects. The term represents the *random effects* for each subject within each group.Finally, represents the independent random error terms, which need to be normally distributed with mean 0 (Davis [2002](#ref-davis2002)). The model then, is a linear combination of terms, and if plotted, it would a straight line.

#### 1.2 Constant correlation

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](#ref-ugrinowitsch2004)). From a practical

#### 1.2 the case of LMEMs 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](#ref-vishwanath2009))

# 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. 2019. “The Reign of the P-Value Is over: What Alternative Analyses Could We Employ to Fill the Power Vacuum?” *Biology Letters* 15 (5): 20190174.

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.”

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.

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*.

1. Department of Biomedical Engineering, University of Arkansas, Fayetteville [↑](#footnote-ref-20)
2. Department of Biomedical Engineering, University of Arkansas, Fayetteville [↑](#footnote-ref-21)
3. Department of Mathematical Sciences, University of Arkansas, Fayetteville [↑](#footnote-ref-22)